# READINGS IN LEARNING AND HUMAN ABILITIES

## EDUCATIONAL PSYCHOLOGY

161

RICHARD E. RIPPLE

8. Education and the Psychology of Individual Differences
*Philip E. Vernon* 103
9. Effects of Accelerating Bright Older Pupils from Second to Fourth Grade
*Herbert J. Klausmeier* and *Richard E. Ripple* 118
10. Sex Differences in the Life Problems and Interests of Adolescents, 1935 and 1957
*Dale B. Harris* 128
11. The Adolescent Subculture and Academic Achievement
*James S. Coleman* 135

## Part III. Teacher Characteristics and Classroom Interactions

12. Some Relationships Between Pupil Behavior and Certain Teacher Characteristics
*David G. Ryans* 152
13. A Study of the Effects of Teacher Knowledge of Pupil Characteristics on Pupil Achievement and Attitudes Towards Classwork
*Kenneth B. Hoyt* 164
14. Teachers' and Clinicians' Attitudes Toward the Behavior Problems of Children: A Reappraisal
*Harry Beilin* 172
15. The Comparative Influence of Punitive and Nonpunitive Teachers upon Children's Concepts of School Misconduct
*Jacob S. Kounin* and *Paul V. Gump* 190
16. Relationships Between Pupil Achievement, Pupil Affect-Need, Teacher Warmth, and Teacher Permissiveness
*C. M. Christensen* 199
17. Group Compliance to Dominative Teacher Influence
*Ned A. Flanders* and *Sulo Havumaki* 207
18. Use of Verbal Reinforcement in Developing Group Discussion Skills
*M. Ray Loree* and *Margaret B. Koch* 225

## Part IV. Learning: Cognitive Abilities and Processes

19. The Act of Discovery
*Jerome S. Bruner* 236
20. Concept Attainment and Adolescent Development
*Lee Yudin* and *Solis L. Kates* 249
21. Vertical and Horizontal Processes in Problem Solving
*Howard H. Kendler* and *Tracy S. Kendler* 258

22. Behaviors During Problem Solving Among Children of Low, Average, and High Intelligence
*Herbert J. Klausmeier* and *Leo J. Loughlin* 279
23. A Definition of Critical Thinking
*Robert H. Ennis* 286
24. The Nature and Nurture of Creative Talent
*Donald W. MacKinnon* 305
25. Caution in Comparing Creativity and IQ
*Richard E. Ripple* and *Frank B. May* 324

**Part V. Learning: Affective Characteristics and Psychomotor Abilities**

26. Attitude Organization in Elementary School Classrooms
*Jules Henry* 330
27. Attitude Change Through Undirected Group Discussion
*K. M. Miller* and *J. B. Biggs* 348
28. Psychological Health and Classroom Functioning: A Study of Dissatisfaction with School Among Adolescents
*Philip W. Jackson* and *Jacob W. Getzels* 355
29. Factors Influencing Change in Children's Self-Concepts
*Hugh V. Perkins* 364
30. A System of the Psychomotor Abilities
*J. P. Guilford* 375
31. The Learning of Motor Skills As Influenced by Knowledge of Mechanical Principles
*Frances M. Colville* 386

**Part VI. Learning: Motivation, Retention, and Transfer**

32. The Motivating Effect of Learning by Directed Discovery
*Bert Y. Kersh* 396
33. Some Motivational Factors in Cooperation and Competition
*Marvin E. Shaw* 405
34. The Use of Advance Organizers in the Learning and Retention of Meaningful Verbal Material
*David P. Ausubel* 420
35. Retention and Transfer in Children of Low, Average, and High Intelligence
*Herbert J. Klausmeier* and *John Check* 428
36. The Transfer Value of Given and Individually Derived Principles
*G. M. Haslerud* and *Shirley Meyers* 435
37. Transfer Effects of Learning a Verbal Generalization
*Susan M. Ervin* 443

## Part VII. Recent Innovations in Teaching–Learning

38. Teaching Machines: A Review
*Charles S. Morrill* 466
39. Education Television
*Walter A. Wittich* 483
40. Laboratories for Foreign Language Teaching
*Milton L. Shane* and *Helen Lacy Shane* 490
41. Team Teaching: An Assessment
*Harold D. Drummond* 497
42. Nongraded Programs: What Progress
*Hugh V. Perkins* 505
43. Administrative Leadership—Key to Core Program Development
*Gordon F. Vars* 511

## Part VIII. Measurement and Evaluation

44. Measurement and the Teacher
*Robert L. Ebel* 528
45. On the Assessment of Academic Achievement
*Henry S. Dyer* 536
46. Testing Programs—Misconceptions, Misuse, Overuse
*Frank B. Womer* 549
47. The Assignment of School Marks
*Marvin D. Glock* and *Jason Millman* 558
48. The Effect of Promotion Policy on Academic Achievement
*G. T. Kowitz* and *C. M. Armstrong* 567
49. Test Anxiety, General Anxiety, and Intellectual Performance
*Irwin G. Sarason* 577

*Index* 589

# Editor's Introduction

The critical necessity for both broad and deep scholarship for the specialist was never as urgent nor as difficult to attain as it is today. Two major circumstances that support the above statement are (1) the vast array of knowledge at hand and (2) the diverse and scattered sources of the published reports by the increasing number of scholars within a special field. The difficulty of both becoming and continuing to be familiar with the literature of the field is especially troublesome for those interested in educational psychology whether they be students, teachers, researchers, or consulting practitioners.

Adequate treatment of the major topics in the field of educational psychology in terms of point of view or school of thought on the one hand, along with supporting evidence (i.e., research) on the other is difficult within a single textbook. However, it is particularly essential that those who are in their early years of study of educational psychology be familiar with various and varying theories, and with the logic, reasoning, and research which justify such theories. *Readings in Learning and Human Abilities,* makes the basic material required for sound scholarship in the field available to any student of educational psychology.

Dr. Richard Ripple, an able and thorough scholar, is also an imaginative, careful, and discriminating researcher. The catholicity of this volume along with the discrimination reflected are telling evidence of Dr. Ripple's high qualification for this assignment. *Readings in Learning and Human Abilities* will increasingly prove invaluable to those who are interested in and concerned with educational psychology.

John Guy Fowlkes

# Preface

Speaking to a group of teachers at Cambridge in 1892, William James had this to say about the relevancy of psychology to education:

> Psychology ought certainly to give the teacher radical help. And yet I confess that, acquainted as I am with the height of some of your expectations, I feel a little anxious lest, at the end of these simple talks of mine, not a few of you may experience some disappointment at the net results . . . I say, moreover, that you make a great, a very great mistake if you think that psychology, being the science of the mind's laws, is something from which you can deduce definite programmes and schemes and methods of instruction for immediate classroom use. Psychology is a science, and teaching is an art, and sciences never generate arts directly out of themselves . . . ( From William James, *Talks to Teachers,* 1892).

While contemporary psychologists would object to defining psychology as "the science of the mind's laws," James' remarks are as appropriate today as they were at the turn of the century. Educational psychology is not (as Coladarci indicates in the third selection in this volume) an ordered catalog of educational recipes that tell us what definite things to say and do when the pupil is before us.

But it is more difficult to define what educational psychology is than to discuss what it is not. Ever since the development of interest in educational psychology at the turn of the century, the field has been in ferment. Questions have been raised as to whether educational psychology should be taught primarily at a scientific–theoretical level or at a technological level; as a basic discipline or as an applied area. Exclusive use of either approach is made at the expense of theoretical issues and orientations or of practical application.

Despite this difference in approach there are some common content areas in educational psychology that are generally agreed upon. It is in the matter of organization of this content that further divergence is apparent. The forty-nine readings included in this volume are arranged into eight sections based on the organization of Herbert J. Klausmeier's *Learning and Human Abilities: Educational Psychology.* Klausmeier's text is designed for introductory courses in educational psychology. It contains practical suggestions for the teacher based on research evidence and set forth in the context of a theory of classroom learning. While research studies in education, psychology, and other areas are reviewed extensively throughout the text, the motivated student will want to examine the original reports of these researchers. Also, material appearing subsequent to the publication of the text and representing further developments in the field will be of interest to the student.

This collection should be particularly useful in colleges with small libraries or in universities with large sections of the introductory course in educational psychology. Often, in these situations, the original sources are not available to students.

The readings related to the material covered in the seventeen chapters of Klausmeier's text are arranged in the following eight sections: I Psychology and Educational Psychology: An Introduction, II Learner Abilities and Characteristics, III Teacher Characteristics and Classroom Interactions, IV Learning: Cognitive Abilities and Processes, V Learning: Affective Characteristics and Psychomotor Abilities, VI Learning: Motivation, Retention, and Transfer, VII Recent Innovations in Teaching-Learning, and VIII Measurement and Evaluation.

Of the forty-nine readings selected, approximately 70 percent were published after 1960 and over 60 percent appear for the first time in a book of readings. The diversity of content covered under the general rubric of educational psychology is reflected in the fact that the readings come from twenty-three different journals. Although the emphasis is on articles reporting empirical research studies, some theoretical articles and some survey articles reviewing research are included. Balance has also been sought through distributing the emphasis on articles dealing with the elementary, secondary, and college levels. In addition to the criteria which may be inferred from the preceding comments, consideration was given to the intended population for which the volume was prepared—college undergraduates in teacher education programs. The intrinsic interest of each article was judged by obtaining reactions from students at Cornell University. No attempt was made to select only the "best" or "classical" articles. The selection of readings was guided by the fact that this volume is intended to accompany Klausmeier's *Learning and Human Abilities: Educational Psychology* with considerations given to recency, balance (in level, content, and type of article), and intrinsic interest of the articles.

Although the *Readings* is a companion to Klausmeier's text, the selections are appropriate for use with any of the current textbooks in educational psychology, and a cross-reference chart correlating the readings with chapters in these texts is on p. xiv. Other features of the volume include a specially-prepared article intended to aid students in reading and interpreting the statistics in the selections, an introduction to each of the eight sections, and an introductory headnote for each reading.

The editor wishes to express his appreciation to the authors and publishers for their permission to reprint the articles included in this volume. Special thanks are due to Mr. Charles Rogers for his work in the initial stages of preparing the manuscript, Mr. Robert O'Reilly for his efforts on the cross-reference chart, and to Mr. George Schuler for his help with the introductory headnotes predicating each article. To the editor's colleagues in the School of Education and to the students in his classes, acknowledgment and gratitude is given for their time and judgment in discussing the selection of readings. For their encouragement and inspirational guidance the editor is deeply indebted to Professors John Guy Fowlkes and Herbert J. Klausmeier of the University of Wisconsin. To Miss Kathy Hillard the editor is grateful for cheerfully bearing up under the mountain of correspondence and other clerical duties surrounding the preparation of this book. And to Miss Millie Pirko the editor wishes to offer his personal gratitude for her competent and always willing assistance in carrying the major burden in the preparation of the manuscript.

R.E.R.

*Ithaca*
*April, 1964*

# Bibliography of Leading Textbooks in Educational Psychology

Bigge, M., and Hunt, M. *Psychological Foundations of Education,* New York: Harper & Row (1962).

Blair, G., Jones, R., and Simpson, R. *Educational Psychology.* New York: The Macmillan Company. Revised (1962).

Cole, L., and Bruce, W. *Educational Psychology.* New York: World Book Company. Revised (1958).

Cronbach, L. *Educational Psychology.* New York: Harcourt, Brace, and World. Revised (1963).

Crow, L., and Crow, A. *Educational Psychology.* New York: American Book Company. Revised (1963).

Frandsen, A. *Educational Psychology: The Principles of Learning in Teaching.* New York: McGraw-Hill Book Company (1961).

Klausmeier, H. *Learning and Human Abilities: Educational Psychology.* New York: Harper & Row (1961).

Lindgren, H. *Educational Psychology in the Classroom.* New York: John Wiley and Sons, Inc. Revised (1962).

McDonald, F. *Educational Psychology.* San Francisco: Wadsworth Publishing Company, Inc. (1959).

Morse, W., and Wingo, G. *Psychology and Teaching.* Chicago: Scott, Foresman, and Company. Revised (1962).

Mouly, G. *Psychology for Effective Teaching.* New York: Holt, Rinehart, and Winston, Inc. (1960).

Pressey, S., Robinson, F., and Horrocks, J. *Psychology in Education.* New York: Harper & Row. Revised (1959).

Sawrey, Jr., and Telford, C. *Educational Psychology.* Boston: Allyn and Bacon, Inc. (1958).

Smith, H. *Psychology in Teaching.* Englewood Cliffs: Prentice-Hall, Inc. Revised (1962).

Thompson, G., Gardner, E., and DiVesta, F. *Educational Psychology.* New York: Appleton-Century-Crofts, Inc. (1959).

Trow, W. *Psychology in Teaching and Learning.* Boston: Houghton Mifflin Company. Revised (1960).

In the cross-reference chart which follows, the selections in this volume are correlated with chapters in the current textbooks in educational psychology listed above.

## Cross-Reference Chart to Leading

| Chapter Numbers | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| igge and Hunt 1962) | 3, 13, 17, 18 | 1 | 2 | 2, 26 | 2, 4, 5, 8, 23, 30, 42, 43, 44, 45, 46, 47 | 2 | 2, 7, 29 | 2, 10, 11, 28 | 26 |
| lair, Jones and Simpson Revised 1962) | 1, 3 | 1, 2, 26, 29 | 2, 7, 29 | 2, 7, 10, 11 | 2, 20, 21 | 2, 8, 9, 22, 33, 42, 43 | 2, 16, 18, 26, 38 | 2, 11, 16, 26, 27, 28 | 19, 20, 21, 22, 23, 24, 25, 34, 35 |
| ole and Bruce (Revised 1963) | 3, 11 | 2, 3, | 2 | 4, 5, 6, 7, 13, 22, 25, 49 | 17, 18 | 1, 16, 26, 33 | 1, 10, 26, 28, 29, 32 | 1, 2, 21, 32 | 1, 2, 18, 31, 38 |
| Cronbach (Revised 1963) | 1, 3 | 1, 2, 3 | 2, 20 | 2, 7, 26, 29 | 43 | 16 | 2, 25, 45, 46 | 2, 4, 5, 8, 9, 13, 20, 22, 23, 25, 42, 45, 46 | 30, 31 |
| Crow and Crow (Revised 1963) | 1 | 1, 17, 18 | 2 | 2 | 2, 7, 29 | 2 | 1, 2, 26 | 2, 4, 5, 6, 20, 25 | 2 |
| Frandsen (1961) | 1, 2, 3 | 1, 18, 21 | 7, 20, 22, 30, 31 | 4, 5, 6, 25, 44, 45, 46, 47 | 8, 9, 11, 27, 42, 43, 48 | 2, 6, 16, 32, 33 | 2, 19, 21, 22, 23, 24, 25, 32 | 18, 19, 23, 38 | 34 |
| Klausmeier (1961) | 1, 2, 3, 4, 7 | 1 | 4, 5, 6, 7, 8, 9, 10, 11, 30 | 3, 12, 13, 14, 15, 16, 17 | 12, 15, 16, 17, 18, 26, 27 | 4, 7, 19, 20 | 4, 19, 21, 22, 23, 24, 25 | 30, 31, 33, 37 | 11, 26, 27, 28 |
| Lindgren (Revised 1962) | 1 | 1, 6, 29, 49 | 2, 7, 20, 30, 31 |  | 17, 26, 32, 33 | 1, 10, 12, 14, 15, 28 | 1, 32, 34, 35, 36, 37 | 2, 18, 19, 21, 22, 24, 32, 34, 35, 36, 37 | 2, 6, 11, 49 |
| McDonald (1959) | 1 | 1, 3 | 1, 3 | 2, 16, 18 | 2, 20 | 2, 19, 20, 21, 22, 23, 24, 25, 32, 34, 35, 36, 37 | 2, 10, 11, 26, 27 | 2, 27, 28, 30, 31, 48 | 18, 38 |
| Morse and Wingo (Revised 1962) | 1, 3 | 1 | 1, 7, 25, 29 | 2, 7 | 2 | 2, 10, 11 | 2, 18, 38 | 2, 6, 19, 21, 22, 23, 24, 32, 49 | 2, 19, 20, 23, 24, 25 27, 30, 31 |
| Mouly (1960) | 3 | 4, 6, 19, 29, 49 | 2 | 2, 30, 31 | 2 | 2, 12, 14, 15, 17, 26 | 2, 4, 5, 7, 20, 25 | 1, 2, 18, 19, 21, 22, 24 | 2 |
| Pressey, Robinson, and Horrocks (Revised 1959) | 1, 2, 3 | 2 | 4, 5, 6, 8, 9 | 10, 11, 28 | 2, 12, 14, 15, 49 | 2, 26 | 2, 7, 16, 18, 35 | 2, 20, 21, 30, 31 | 2, 34, 35 36, 37 |
| Sawrey and Telford (1958) | 1, 2, 3, 11, 16, 28, 29, 32, 33 | 1, 2, 19, 21, 22, 23, 24, 25 | 2, 7, 18 | 2, 34, 35 | 35, 36, 37 | 13, 20, 44, 45, 46, 47 | 2 | 2 | 2, 4, 5, 6, 8, 9, 22, 25, 45, 46 |
| Smith (Revised 1962) | 1, 3 | 2, 20, 29 | 2, 49 | 2, 17, 18, 26 | 2, 11, 27 | 4, 5, 6, 7, 13, 22, 25, 45, 46 | 2, 4, 5, 8, 9, 20, 22, 25, 42, 43 | 2, 10, 11, 16, 32, 33 | 2, 20, 34, 35, 36, 37 |
| Thompson, Gardner, and DiVesta (1959) | 1, 2, 7 | 2 | 2, 28 | 2, 13, 44, 45, 46, 47, 49 | 13, 44, 45, 46, 47, 49 | 4, 5, 6, 13, 22, 25, 45, 46, 49 | 10, 46 | 2, 11, 12, 14 | 2, 26 |
| Trow (1960) | 1, 3 | 2, 27, 28 | 2, 7 | 2, 16, 33 | 2 | 2, 13, 22, 44, 45, 46, 47 | 2, 4, 5, 6, 8, 9, 13, 25, 45, 46, 48, 49 | 2 | 2, 29 |

| 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|
| 2, 20, 21 | 1, 2, 21 | 2, 19, 32 | 1, 2, 18, 21, 23, 38 | 2 | 2, 16, 31, 32, 33, 34, 35, 36 37 | 14, 24, 25 | 2, 32 | 2, 6, 19, 21, 22, 23, 24, 25, 32, 49 | 12, 15, 16, 17, 18, 27, 32 |
| 32, 35, 36, 37 | 2, 12, 15, 16, 17, 18, 26 |  | 26 | 12, 14, 15 | 2, 26 |  | 2, 6, 24, 25, 44, 45, 46 | 2, 13, 44, 45, 46, 47 | 47, 48, 49 |
| 1, 2, 18, 20, 31, 32 | 2, 32 | 19, 20, 21, 22, 23, 34, 35, 36, 37 | 24, 25 | 3, 29 | 2, 13, 44, 45, 46, 47 | 12, 15 |  |  |  |
| 35, 36, 37 | 19, 21, 22, 23, 24, 25, 32, 34, 35 | 38, 39, 40, 41 | 2, 10, 11, 26, 27 | 14, 18, 28 | 2, 12, 15, 16, 17, 18 20, 26, 32, 33 | 2, 13, 44, 45, 46, 47, 48 | 6, 14, 24, 49 | 1 |  |
| 10 | 2, 8, 9, 42, 43 | 2, 18, 21, 32, 38 | 2, 10, 11, 16, 27, 28, 33, 36 |  | 30, 31, 37 | 1, 19, 21, 22, 23, 24, 25 | 35, 36, 37 | 2, 13, 44, 45, 46, 47, 48 | 2, 3, 44, 45, 46, 47 |
| 2, 34, 35, 36, 37 | 2 | 6, 26, 28, 49 | 14 | 2, 12, 15, 17, 18, 26 29 | 2, 13, 34, 35, 44, 45, 46, 47 |  |  |  |  |
| 6, 16, 26, 28, 29, 49 | 2, 16, 19, 32, 33 | 2, 34, 35, 36, 37 | 2, 8, 9, 13, 22, 42, 43, 48 | 2, 38, 39, 40, 41, 42, 43 | 2, 44, 45, 46, 47, 48, 49 | 2, 44, 45, 46, 47, 48, 49 | 47 |  |  |
| 1, 2, 12, 17, 32, 38 | 2, 17, 23, 24, 25, 27, 32, 39 | 15 | 2, 17, 18, 26, 27 | 2, 25, 44, 45, 46, 47, 48 | 2, 4, 5, 6, 8, 9, 13, 25, 45, 46 | 22, 42, 43 | 10 | 3 |  |
| 19, 21, 22, 24, 25, 36, 37 | 2, 7, 20, 22, 42, 43 | 29 | 2, 12, 14, 15, 17, 18, 26, 32, 33 | 2, 8, 9, 44 | 2, 25, 44, 45, 46, 47, 48 | 2, 4, 5, 6, 8, 9, 13, 45, 46, 47 | 2, 11 | 3 |  |
| 2, 34, 35, 36, 37 | 2, 4, 5, 7, 8, 9, 20, 22, 35 | 2, 16, 32, 33, 48 | 2, 11, 17, 18, 26, 27, 42, 43 | 12, 14, 15, 26, 28 | 2, 12, 14 | 2, 3, 4, 13, 44, 45, 46 |  |  |  |
| 2, 11, 32, 33 | 2, 34, 35, 36, 37 | 19, 21, 22, 23, 24, 25 | 2, 27, 28 | 13, 44, 45, 46, 47, 48 | 2, 8, 9, 22, 42, 43, 48 | 2, 10 | 14 | 3 |  |
| 2, 16, 17, 18, 26, 27, 33 | 1, 2, 3, 32, 48 |  | 2, 13, 25, 44, 45, 46, 47 | 22, 42, 43 |  | 1, 10 | 19, 23, 24, 25, 29 |  |  |
| 10 | 2, 30, 31 | 2, 27 | 2, 17, 18 26 | 12, 14, 15, 49 | 1 |  |  |  |  |
| 2, 11, 17, 19, 21, 22, 23, 24, 25, 32, 34, 35 |  | 2, 13, 44, 45, 46, 47, 48 | 2, 14, 28 | 2, 12, 14, 15, 28 | 1 |  |  |  |  |
| **7** | 2, 7, 20, 23, 42, 43 | 2, 7, 17, 32, 33, 48 | 7, 16, 34, 35, 36, 37 | 2, 19, 21, 22, 23, 24, 25, 32 | 2, 20 | 1, 2, 11, 27, 28 | 30, 31 | 26, 28 | 29 |
| 2, 18, 30, 31, 38 | 2, 34, 35, 36, 37 | 2, 19, 20, 21, 22, 23, 24, 25, 32 | 2, 12, 14, 15, 26 | 2, 17, 18, 26, 42, 43 | 2 |  |  |  |  |

# Cross-Reference Chart to Leading Textbooks in Educational Psychology *(Continued)*

| Chapter Numbers | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
|---|---|---|---|---|---|---|---|---|---|---|
| Bigge and Hunt (1962) | | | | | | | | | | |
| Blair, Jones and Simpson (Revised 1962) | 48 | 3 | | | | | | | | |
| Cole and Bruce (Revised 1963) | | | | | | | | | | |
| Cronbach (Revised 1963) | | | | | | | | | | |
| Crow and Crow (Revised 1963) | 2, 44, 45, 46, 47 | 41 | | 19, 24 | 24 | | | 10 | 26 | 14 |
| Frandsen (1961) | | | | | | | | | | |
| Klausmeier (1961) | | | | | | | | | | |
| Lindgren (Revised 1962) | | | | | | | | | | |
| McDonald (1959) | | | | | | | | | | |
| Morse and Wingo (Revised 1962) | | | | | | | | | | |
| Mouly (1960) | | | | | | | | | | |
| Pressey, Robinson, and Horrocks (Revised 1959) | | | | | | | | | | |
| Sawrey and Telford (1958) | | | | | | | | | | |
| Smith (Revised 1962) | | | | | | | | | | |
| Thompson, Gardner, and DiVesta (1959) | 2 | 19, 32 | 2, 12, 15, 16, 17, 18, 26 | | | | | | | |
| Trow (1960) | | | | | | | | | | |

# An Introduction to Statistical Analysis and Terminology — *Jason Millman*

ONE OF THE characteristics of a social science such as educational psychology is that it reports in its professional literature the results of its researches quantitatively. Phenomena are measured and counted, but only a portion of such data is presented. Averages and other statistical indexes and techniques are used to reduce the data to forms like the ones found in the readings in this book. A major purpose of the present article will be to discuss these indexes and techniques so that the readings can be followed with greater understanding.

## Measures of Central Tendency

The most frequently used index in these readings is the *mean*. The mean of 3, 4, and 8 is 5; that is, 15 (the sum of the numbers) divided by 3 (the number of values added). Characteristic of the mean is that the *deviations* of the numbers above it balance the deviations of the numbers below it. Thus in the example above, the deviation of 8 from 5, or 3, is equal to the sum of the deviations of 3 and 4 from the mean. For this reason, the mean is considered as that value which is at the center of gravity of the numbers.

The *median*, another widely used type of average, is that value above and below which an equal number of values lie. The median of the numbers 3, 4, and 8 is thus 4.

It is possible to compute both a mean and a median for any set of numbers, regardless of what the numbers signify. In the readings in this volume, means or medians represent test scores, correlation coefficients, ages, number of trials, and so forth.

The means or medians of two or more sets of data are frequently compared. For example, Ausubel (34, Table 1)[1] reports the mean retention score of experimental and control groups to be 16.7 and 14.1 respectively. The experimental group has, on the average, greater retention than the control group. Although every member in the first does not necessarily have a greater retention score than every member in the second group, there is an average difference between the groups of 2.6 points.

## Measures of Variability

The *range* is a single numbered index which indicates the spread of values from one extreme to the other. For example, a day's temperature range might be 15°, from 50° to 65°. Since the calculation of the range depends upon only two numbers, and is therefore highly unstable, most investigators prefer to use *standard deviation* and *variance* as measures of the spread or dispersion of a set of numbers. These statistics depend upon all the values.

One way to compute a standard deviation and variance is to first calculate the square of the deviations of each number from the mean of all numbers considered. The variance is the mean of these squared deviations; the standard deviation is equal to the square root of the variance. Hence the greater the deviations, the greater the spread of the numbers, and the greater the value of the standard deviation and variance.

In the Ausubel study cited previously, the standard deviation of retention scores for the experimental group was 5.8 and for the control group 5.4. Thus, the spread of scores for the experimental group was greater than it was for the control group. To put it another way, the experimental group had more heterogeneous retention scores than the control group. In this example, the experimental group had both the higher mean and the greater standard deviation. This is not always the case, however, and the group with the higher mean may have the smaller standard deviation.

One characteristic of the standard deviation which aids in its interpretation is that, for most sets of data, 55 to 75% of the numbers are within one standard deviation of the mean. Thus for Ausubel's experimental group, which had a mean retention score of 16.7 and a standard deviation of 5.8, approximately two-thirds of the subjects can be expected to have retention scores between 10.9 and 22.5.

Sometimes investigators wish to compare the performance of an individual on several *variables* (attributes possessing two or more values), each variable having a different mean. A crude procedure is to indicate on which variables the individual was above the mean, and on which he was below. A more precise procedure is to determine how many standard deviations above or below its respective mean is each score. Scores based on such standard deviation units are called *standard scores.* (For example, see 12, 28, 47.)

---

[1] Numbers by themselves, enclosed in parentheses, refer to the numbers of the readings in this book. Thus, here the reference is to reading number 34.

### Measures of Relationship

The primary aim of most of the research articles in this volume is to determine the existence and magnitude of relationships between variables. Some variables such as sex have only two values (male or female) and are called *dichotomies.* Individuals can be divided into two groups on the basis of which value they possess on a dichotomous variable. The relation between two variables, one of which is a dichotomy, can be then provided by a comparison of each group's mean value on the second variable. For example, in the study noted previously, Ausubel (34) reported a relation between type of group (experimental or control) and retention scores. That is, there was a difference between the mean retention scores for the two groups. Had there been virtually no overlap in the distribution of retention scores for the two groups, that is, had almost all of the highest scores been made by subjects in the experimental group, the relation would have been close to perfect.

An index which indicates the amount of straight-line or linear relationship between two variables is the *Pearson product-moment correlation, r.* To compute a correlation coefficient, two values are needed for each of several subjects. If the subjects who have high values on one variable have also high values on the second variable, the correlation coefficient is positive. If the tendency is the reverse, that is, if high values on the first variable are associated with low values on the second, the coefficient is negative. The closer to plus or minus 1.00, the stronger the relation. A correlation coefficient of either plus or minus 1.00 indicates a perfect relationship and therefore perfect prediction. A correlation coefficient of .00 represents no relation. For example, the height of a group of subjects, when measured in feet, would correlate 1.00 with their height when measured in inches, since an individual's height in feet could be perfectly predicted from knowledge of his height in inches. High-school and colleges grades correlate approximately .50. Shoe size and intelligence would be expected to correlate approximately zero.

When Feldhusen and Klausmeier (6, Table 4) correlated anxiety and reading scores for students of average intelligence, they obtained a correlation of $-.31$. This negative correlation indicates a tendency for the *more* anxious students to have *lower* reading scores. The corresponding correlation for students of high intelligence was +.15. In this case, there was a slight tendency for the more anxious students to have better reading scores.

Although correlation coefficients may look like percentages, it is incorrect to interpret them as such. In terms of predictive efficiency, a correlation of 1.00 is more than twice as good as a correlation of .50, which in turn is more than twice as good as one of .25.

The *biserial correlation* (5) is also a product-moment correlation and may may be interpreted similarly. The *rank order correlation, rho,* (10, 14, 16) is the product-moment correlation $r$ between the ranks of the values rather than between the actual values themselves. A slightly different measure of relation is the *tau coefficient* (37). It too has limits of plus and minus 1.00, and in the special case when both variables are a dichotomy, the tau coefficient is numerically equal to $r$. *Epsilon* (49) is a measure of curvilinear relation. For example, over the entire life span of a person, age and strength, if graphed, would form a curve. The epsilon correlation coefficient between age and strength would therefore be much higher than a measure of straight-line relation such as $r$. Only in the special case when the curve which best fits the data is a straight line, would epsilon be as low as $r$.

Often students mistakenly believe that the size of a correlation depends upon the number of cases used. If this were true, one could obtain a near perfect relation or prediction between two variables by measuring a large number of people. In fact, increasing the number of cases does not necessarily result in higher correlations, although it does tend to bring the computed correlation closer to the true correlation.

Throughout this volume, *reliability* of the measuring instruments being employed is reported. An instrument possesses reliability to the extent that the same individuals have the same relative standing on two testings with the same instrument or on two equivalent instruments. This is expressed numerically as a correlation coefficient. Most carefully made standardized tests have reliability coefficients of over .80; however, the reliability coefficients of short, experimental measures of personality constructs are often much lower.

When a large number of variables are correlated with each other, *factor analysis* (12) is sometimes used to reduce the data. Instead of indicating the correlation of each variable with each other one, factor analysis provides correlation coefficients between each variable and each of a fewer number of hypothetical variables, called *factors*. Each factor represents what is common to several variables. The correlations between the variables and the factors are called *factor loadings*. The first factor loading in Table 1 of Ryans' article (12) is .07. There is a .07 correlation between pupil behavior on the apathetic-alert dimension and a hypothetical variable, Factor A.

## Tests of Statistical Significance

A frequently asked question is whether chance alone can account for the size of certain statistical indexes or for the magnitude of the difference between corresponding statistical indexes of two or more groups.

If a coin were tossed in the air ten times, chance alone could account for it landing heads seven times. The question is: how many of the tosses must be heads, or how large must be the difference between the number of heads and tails, before the likelihood or *probability* of chance alone operating is so low that one is led to believe the coin or the process of flipping is biased?

A researcher may ask whether the difference of two means is sufficiently large so that chance alone could no longer be expected to account for the difference; or, he may want to know whether an observed correlation of .31 is sufficiently greater than the chance correlation of .00 to conclude that the two variables are truly related.

To answer such questions, the researcher determines the probability that the observed statistical values (or ones more extreme) could occur *if* chance alone were operating. If this probability is low—traditionally five chances in a hundred (.05), or less—he rejects the hypothesis that chance alone accounts for the results and considers them *statistically significant.* The lower the probability, the less likely it is that the results could be caused solely by chance, and the more statistically significant are the findings. If the probability of results in only one direction are considered, the test is known as a *one-tailed test* (34, 37).

Such statements as, "significant beyond the .05 level," "significant at the 5% level," "reliable beyond the .05 level," "significant beyond the .05 level of confidence," "probability less than five chances in one hundred," and "$p<.05$" are synonymous. If the probability of obtaining results if only chance were operating were greater than .05, the results would not be considered statistically significant.

Whether a particular result is statistically significant depends upon the size of the sample. While 70% heads landing in ten coin tosses may not be considered unusual, even 55% heads in 10,000 tosses could virtually never happen by chance alone. Thus, it is possible for a small difference (55% vs. 50% in the example just given), or a low correlation computed on a large sample to be statistically significant, but not be very large in an absolute sense. The distinction between statistical significance and practical significance, or importance must be remembered.

If, to vary the original example, ten coins were welded to each other with the same side up, a toss of the ten coins at once would have to result in either all heads or all tails. In this situation, an event of 10 out of 10 heads would not be surprising, whereas if the ten coins were tossed independently, 10 out of 10 heads would be highly significant. Technically, then, it is the number of independent cases, called *degrees of freedom*, rather than the actual size of the sample, upon which significance most directly depends.

Depending upon the statistics and the type of data with which he is working, as well as the assumptions he is willing to make, the investigator can often determine both theoretical distributions of possible results and the probabilities of those results occurring, were chance alone operating. The statistical test that the investigator is using is often named after the theoretical distribution appropriate for his set of circumstances. The $t$, normal, $F$, and chi-square distributions are most frequently employed. This volume includes selections using *Fisher's t* and the *normal-curve critical-ratio test* (when testing the significance of the difference between two means), the *F test* (when testing the significance of the difference between two or more means), and the *chi-square test* (when testing the significance of the difference between percentages or frequencies). For a given number of degrees of freedom, the larger the value of the $t$, critical-ratio, $F$, or chi-square statistic, the more significant the results.

| | | Relevant | None | Irrelevant |
|---|---|---|---|---|
| Age | 4 | | | |
| | 7 | | | |

FIG. 1

Other statistical tests included in this volume are the *Mann-Whitney test* (37), which indicates essentially the significance of the difference between two medians: the *Tukey gap test* (6), the *Dunnett procedure* (9), and the *Duncan range test* (20)—all of which test the significance of the difference between the means of specific groups when several groups are being compared. *Analysis of variance* makes use of the $F$ distribution and can test in one process the significance of the difference between several means. *Analysis of covariance* (16, 18, 35) is similar to the analysis of variance, except that the means to be compared are first adjusted to account for differences between the groups on one or more uncontrolled variables.

To illustrate additional statistical terms used in this book, consider the situation (21) where children are grouped on two independent variables, as shown in Figure 1. This diagram is a 2 × 3 *table* (15) or, if the six *cells* contained the number of children involved, a 2 × 3 *contingency table* (32). The analysis of variance technique, which uses an $F$ test, can test the significance of the difference between mean scores of the two age groups on the dependent variable (in this case a measure of learning), the significance of the difference between the mean scores of students receiving the three types of verbal condition, and the *interaction* between age and type of verbal condition received on the learning measure. An interaction would exist if the superiority, on the learning measure, of the 7-year-old children over the 4-year-old children was different for groups of children receiving different verbalization conditions. Figure 7 in Kendler and Kendler's article (21) clearly shows this interaction. The younger children profited more from one type of verbal condition, and the older children from another. Since both some younger and some older children received each type of verbal condition, this is an example of a 2 × 3 *factorial design* (13, 16).

## Statistical Symbols and Abbreviations

| *Symbol or Abbreviation* | *Meaning* | *Selection Where Used* |
|---|---|---|
| C.R. | critical ratio | (5) |
| $df$ | degrees of freedom | several |
| M | mean | several |
| Mdn. | median | (5) |
| $MS$ | mean square; an intermediate value calculated during an analysis of variance | (6, 16) |
| $N$ | number of cases | several |
| $ns$ | not significant, statistically | (22, 28, 37) |
| $p$ | probability | several |
| $r$ | Pearson product-moment correlation coefficient; sometimes a reliability index | several |
| $r_{\text{bis}}$ | biserial correlation | (5) |
| s | standard deviation | (28) |
| $SD$ | standard deviation | several |
| $SS$ | sum of squares; an intermediate value calculated during an analysis of variance | (20) |
| $\bar{x}$ | mean | (28, 36) |
| x or X | by; examples: 2 × 3 is read two by three; A × B is read A by B or the interaction of variables A and B | several |
| $z$ | a value to which correlations are transformed before being averaged | (29) |
| $z$ | the end calculation of a critical ratio test | (29) |
| $<$ | less than | several |
| $\leq$ | less than or equal to | (20) |
| $>$ | greater than | (22) |
| $\chi^2$ | chi-square | (15, 17, 33, 37) |
| $\partial$ | a corruption of the more frequently used symbol for standard deviation, $\sigma$ | (36) |

# Part I

## *Psychology and Educational Psychology:*

## *An Introduction*

WHAT IS the relationship among psychology, educational psychology, and education? Or, what should it be? One popular view is that educational psychology should serve as an engineering manual that bridges the gap between psychological theory and research on the one hand, and the practice of teaching on the other. But, as James puts it, "psychology is a science and teaching is an art; and sciences never generate arts directly out of themselves." How, then, can educational psychology bridge this gap?

The three readings contained in Part I: Psychology and Educational Psychology: An Introduction, bear on these questions and serve as an introduction to the present book of readings. In the first selection "The Place of the Person in the New World of the Behavioral Sciences," Carl Rogers discusses the breadth, depth, and extent of advances in psychology and the behavioral sciences. Professor in the departments of Psychology and Psychiatry at the University of Wisconsin, Rogers presents examples of the increased ability of psychology to understand, predict, and control behavior.

Goodwin Watson, of Teachers College, Columbia University, discusses in the second selection the problems of educators who wish to apply psychology in their professional work. He presents fifty propositions important for educators; these are based on "solid psychological ground and not on shifting sands."

In the third and final selection, Arthur Coladarci deals directly with the relevancy of educational psychology. Coladarci, professor of education and psychology at Stanford University, sees the contributions of educational psychology as falling into two related categories—as a body of information and an arena of research activity, and as a help to teachers and administrators in acquiring the attitudes and skills necessary to intelligent hypothesizing and to the testing of hypotheses.

# 1

# The Place of the Person in the New World of the Behavioral Sciences — *Carl R. Rogers*

*The ultimate goal of science is understanding: more proximate aims include the prediction and control of events. Citing advances in psychology and the behavioral sciences, the author cautions in the present article that this progress will create profound problems of decision in the public domain. Values and purposes in scientific research, in educational psychology, and in the classroom are subjectively chosen; they can never be tested nor even formulated by the particular scientific endeavor to which they give birth and meaning. This view of the "great paradox of the behavioral sciences" should stimulate further animated discussion.*

THE SCIENCE of psychology, in spite of its immaturities and its brashness, has advanced mightily in recent decades. From a concern with observation and measurement it has increasingly moved toward becoming an "if-then" science. By this I mean it has become more concerned with the discernment and discovery of lawful relationships such that *if* certain conditions exist, *then* certain behaviors will predictably follow. It is rapidly increasing the number of areas or situations in which it may be said that if certain describable, measurable conditions are present or are established, then predictable, definable behaviors are learned or produced.

Now in one sense every educated person is aware of this. But it seems to me that few are aware of the breadth, depth, and extent of these advances in

Reprinted from *Personnel and Guidance Journal*, February, 1961, pp. 442–451, by permission of the author and publisher.

psychology and the behavioral sciences. And still fewer seem to be aware of the profound social, political, ethical, and philosophical problems posed by these advances. I would like to focus on some of the implications of these advances.

Let me venture first to review a few selected examples of what I mean by the increased ability of psychology to understand and predict or control behavior. I have chosen them to illustrate the wide range of behaviors involved. I shall summarize and greatly simplify each of the illustrations, with only a suggestion of the evidence which exists. As a general statement I may say that each illustration I will give is supported by reasonably rigorous and adequate research, though like all scientific findings, each is open to modification or correction through more exact or imaginative future studies.

What, then, are some of the behaviors or learnings for which we now know how to supply the antecedent conditions? I would stress that we know how to produce these effects in the same way, though not with the same exactitude, that the physicist knows how to set up the conditions under which given substances will go through a process of atomic fission or fusion. They are instances of what we know how to achieve or accomplish.

*We know how to set up the conditions under which many individuals will report as true, judgments which are contrary to the evidence of their senses.* They will, for example report that Figure A covers a larger area than Figure B, when the evidence of their senses *plainly* indicates that the reverse is true. Experiments by Asch (*1*) later refined and improved by Crutchfield (*4*) show that when a person is led to believe that everyone else in the group sees A as larger than B, then he has a strong tendency to go along with this judgment and in many instances does so with a real belief in his false report.

Not only can we predict that a certain percentage of individuals will thus yield and disbelieve their own senses, but Crutchfield has determined the personality attributes of those who will do so and by selection procedures would be able to choose a group who would almost uniformly give in to these pressures for conformity.

*We know how to change the opinions of an individual in a selected direction, without his ever becoming aware of the stimuli which changed his opinion.* A static, expressionless portrait of a man was flashed on a screen by Smith, Spence, and Klein (*17*). They requested their subjects to note how the expression of the picture changed. Then they intermittently flashed the word "angry" on the screen, at exposures so brief that the subjects were consciously completely unaware of having seen the word. They tended, however, to see the face as becoming more angry. When the word "happy" was flashed on the screen in similar fashion, the viewers tended to see the face as becoming more happy. Thus they were clearly influenced by stimuli which registered at a subliminal level, stimuli of which the individual was not, and could not be, aware.

*We can predict, from the way individuals perceive the movement of a spot of light in a dark room, whether they tend to be prejudiced or unprejudiced.* There has been much study of ethnocentrism, the tendency toward a pervasive and rigid distinction between ingroups and outgroups, with hostility toward outgroups, and a submissive attitude toward, and belief in the rightness of, ingroups. One of the theories which has developed is that the more ethnocentric person is unable to tolerate ambiguity or uncertainty in a situation. Operating on this theory Block and Block (*3*) had subjects report on the degree of movement they perceived in a dim spot of light in a completely dark room. (Actually no movement occurs, but almost all individuals perceive movement in this situation.) They also gave these same subjects a test of ethnocentrism. It was found, as predicted, that those who, in successive trials, quickly established a norm for the amount of movement they perceived, tended to be more ethnocentric than those whose estimates of movement continued to show variety.

This study was repeated, with slight variation, in Australia (*18*), and the findings were confirmed and enlarged. It was found that the more ethnocentric individuals were less able to tolerate ambiguity, and saw less movement than the unprejudiced. They also were more dependent on others and when making their estimates in the company of another person, tended to conform to the judgment of that person.

Hence it is not too much to say that by studying the way the individual perceives the movement of a dim light in a dark room, we can tell a good deal about the degree to which he is a rigid, prejudiced, ethnocentric person.

*We know the attitudes which, if provided by a counselor or a therapist, will be predictably followed by certain constructive personality and behavior changes in the client.* Studies we have completed in recent years in the field of psychotherapy (*11–13, 19*) justify this statement. The findings from these studies may be very briefly summarized in the following way.

If the therapist provides a relationship in which he is (a) genuine, internally consistent; (b) acceptant, prizing the client as a person of worth; (c) empathically understanding of the client's private world of feelings and attitudes; then certain changes occur in the client. Some of these changes are: the client becomes (a) more realistic in his self-perceptions; (b) more confident and self-directing; (c) more positively valued by himself; (d) less likely to repress elements of his experience; (e) more mature, socialized, and adaptive in his behavior; (f) less upset by stress and quicker to recover from it; (g) more like the healthy, integrated, well-functioning person in his personality structure. These changes do not occur in a control group and appear to be definitely associated with the client's being in a relationship with these qualities.

*We know how to provide animals with a most satisfying experience consisting entirely of electrical stimulation.* Olds (6) has found that he can implant tiny electrodes in the septal area of the brain of laboratory rats. When one of these animals presses a bar in his cage, it causes a minute current to pass through these electrodes. This appears to be such a rewarding experience that the animal goes into an orgy of bar pressing, often until he is exhausted. Whatever the subjective nature of the experience it seems to be so satisfying that the animal prefers it to any other activity. I will not speculate as to whether this procedure might be applied to human beings, nor what, in this case, its consequence would be.

*We know how to provide psychological conditions which will produce vivid hallucinations and other abnormal reactions in the thoroughly normal individual in the waking state.* This knowledge came about as the unexpected by-product of research at McGill University (2). It was discovered that if all channels of sensory stimulation are cut off or muffled, abnormal reactions follow. If healthy subjects lie motionless to reduce kinaesthetic stimuli, with eyes shielded by translucent goggles which do not permit perception, with hearing largely stifled by foam rubber pillows as well as by being in a quiet cubicle, and with tactile sensations reduced by cuffs over the hands, then hallucinations and bizarre ideation bearing some resemblance to that of the psychotic occur within a relatively short time in most subjects. What the results would be if the sensory stifling were continued longer is not known because the experience seemed so potentially dangerous that the investigators were reluctant to continue it.

I hope that these few illustrations will have given some concrete meaning to the statement that the behavioral sciences are making strides in the understanding, prediction, and control of behavior. In important ways we know how to select individuals who will exhibit certain behaviors; to establish conditions in groups which will lead to various predictable group behaviors; to establish conditions which, in an individual, will lead to specified behavioral results; and in animals our ability to understand, predict, and control goes even further, possibly foreshadowing future steps in relation to man.

If your reaction is the same as mine then you will have found that this picture I have given has its frightening as well as its strongly positive aspects. With all the immaturity of this young science, and its vast ignorance, even its present state of knowledge contains awesome possibilities. Perhaps it makes clear the reason why Robert Oppenheimer, one of the most gifted of our natural scientists, looks out from his own domain of physics, and out of the experiences in that field voices a warning. He says that there are some similarities between physics and psychology, and one of these similarities "is the extent to which our progress will create profound problems of decision in the public domain. The physicists have been quite noisy about their contributions in the last decade. The time may well come—as psychology acquires a sound objective corpus of knowledge about human behavior and feeling—when the powers of control thus available will pose far graver problems than any the physicists have posed" (*7*).

Among behavioral scientists it seems to be largely taken for granted that the finding of such science will be used in the prediction and control of human behavior. Yet most psychologists and other such scientists have given little thought to what this would mean.

I should like to try to present, as well as I can, a simplified picture of the cultural pattern which emerges if we endeavor to shape human life in terms of the behavorial sciences. This is one of two possible directions I wish to consider.

There is first of all the recognition, almost the assumption, that scientific knowledge is the power to manipulate. Dr. B. F. Skinner of Harvard says: "We must accept the fact that some kind of control of human affairs is inevitable. We cannot use good sense in human affairs unless someone engages in the design and construction of environmental conditions which affect the behavior of men" (*14*).

Let us look at some of the elements which are involved in the concept of the control of human behavior as mediated by the behavioral sciences. What would be the steps in the process by which a society might organize itself so as to formulate human life in terms of the science of man?

First would come the selection of goals. In a recent paper (14) Dr. Skinner suggests that one possible goal to be assigned to the behavior technology is this: "Let man be happy, informed, skillful, well-behaved, and productive." In his book, *Walden Two* (*16*), where he can use the guise of fiction to express his views, he becomes more expansive. His hero says, "Well, what do you say to the design of personalities? Would that interest you? The control of temperament? Give me the specifications, and I'll give you the man! What do you say to the control of motivation, building the interests which will make men most productive and most successful? Does that seem to you fantastic? Yet some of the techniques are available, and more can be worked out experimentally. Think of the possibilities . . . Let us control the lives of our children and see what we can make of them."

What Skinner is essentially saying here is that the current knowledge in the behavior sciences, plus that which the future will bring, will enable us to specify, to a degree which today would seem incredible, the kind of behavioral and personality results which we wish to achieve.

The second element in this process would be one which is familiar to every scientist who has worked in the field of applied science. Given the purpose, the goal, we proceed by the method of science—by controlled experimentation—to discover the means to these ends. The method of science is self-correcting in thus arriving at increasingly effective ways of achieving the purpose we have selected.

The third element in the control of human behavior through the behavioral sciences involves the question of power. As the conditions or methods are discovered by which to achieve our goal, some person or group obtains the power to establish those conditions or use those methods. There has been too little recognition of the problem involved in this. To hope that the power which is being made available by the behavioral sciences will be exercised by the scientists, or by a benevolent group, seems to me a hope little supported by either recent or distant history.

It seems far more likely that behavioral scientists, holding their present attitudes, will be in the position of the German rocket scientists specializing in guided missiles. First they worked devotedly for Hitler to destroy Russia and the United States. Now, depending on who captured them, they work devotedly for Russia in the interest of destroying the United States, or devotedly for the United States in the interest of destroying Russia. If behavioral scientists are concerned solely with advancing their science, it seems most probable that they will serve the purposes of whatever individual or group has the power.

But this is, in a sense, a digression. The main point of this view is that some person or group will have and use the power to put into effect the methods which have been discovered for achieving the desired goal.

The fourth step in this process whereby a society might formulate its life in terms of the behavioral sciences is the exposure of individuals to the methods and conditions mentioned. As individuals are exposed to the prescribed conditions this leads, with a high degree of probability, to the behavior which had been desired. Men then become productive, if that has been the goal, or submissive, or whatever it has been decided to make them.

To give something of the flavor of this aspect of the process as seen by one of its advocates, let me again quote the hero of *Walden Two*. "Now that we *know* how positive reinforcement works, and why negative doesn't," he says, commenting on the method he is advocating, "we can be more deliberate and hence more successful, in our cultural design. We can achieve a sort of control under which the controlled, though they are following a code much more scrupulously than was ever the case under the old system, nevertheless *feel free*. They are doing what they want to do, not what they are forced to do. That's the source of the tremendous power of positive reinforcement—there's no restraint and no revolt. By a careful cultural design, we control not the final behavior, but the *inclination* to behave—the motives, the desires, the wishes. The curious thing is that in that case *the question of freedom never arises*" (*16, p. 218*).

## The Picture and Its Implications

Let me see if I can sum up very briefly the picture of the impact of the behavioral sciences upon the individual and upon society, as this impact is explicitly seen by Dr. Skinner and implied in the attitudes and work of many, perhaps most, behavioral scientists. Behavioral science is clearly moving forward; the increasing power for control which it gives will be held by some one or some group; such an individual or group will surely choose the purposes or goals to be achieved; and most of us will then be increasingly controlled by means so subtle we will not even be aware of them as controls. Thus whether a council of wise psychologists (if this is not a contradiction in terms) or a Stalin or a Big Brother has the power, and whether the goal is happiness, or productivity, or resolution of the Oedipus complex, or submission, or love of Big Brother, we will inevitably find ourselves moving toward the chosen goal, and probably thinking that we ourselves desire it. Thus if this line of reasoning is correct, it appears that some form of completely controlled society—a Walden Two or a 1984—is coming. The fact that it would surely arrive piecemeal rather than all at once, does not greatly change the fundamental issues. Man and his behavior would become a planned product of a scientific society.

You may well ask, "But what about individual freedom? What about the democratic concepts of the rights of the individual?" Here too Dr. Skinner is quite specific. He says quite bluntly, "The hypothesis that man is not free is essential to the application of scientific method to the study of human behavior. The free inner man who is held responsible for his behavior . . . is only a prescientific substitute for the kinds of causes which are discovered in the course of scientific analysis. All these alternative causes lie *outside* the individual" (*15, p. 447*).

I have endeavored, up to this point, to give an objective picture of some of the developments in the behavioral sciences and an objective picture of the kind of society which might emerge out of those developments. I do however have strong personal reactions to the kind of world I have been describing, a world which Skinner explicitly (and many another scientist implicitly) expects and hopes for in the future. To me this kind of world would destroy the human person as I have come to know him in the deepest moments of psychotherapy. In such moments I am in relationship with a person who is spontaneous, who is responsibly free, that is, aware of his freedom to choose whom he will be and aware also of the consequences of his choice. To believe, as Skinner holds, that all this is an illusion and that spontaneity, freedom, responsibility, and choice have no real existence would be impossible for me.

I feel that to the limit of my ability I have played my part in advancing the behavioral sciences, but if the result of my efforts and those of others is that man becomes a robot, created and controlled by a science of his own making, then I am very unhappy indeed. If the good life of the future consists in so conditioning individuals through the control of their environment and through the control of the rewards they receive, that they will be inexorably productive, well-behaved, happy, or whatever, I want none of it. To me this is a pseudo-form of the good life which includes everything save that which makes it good.

And so I ask myself, is there any flaw in the logic of this development? Is there any alternative view as to what the behavioral sciences might mean to the individual and to society? It seems to me that I perceive such a flaw and that I can conceive of an alternative view. These I would like to set before you.

### Ends and Values in Relation to Science

It seems to me that the view I have presented rests upon a faulty perception of goals and values in their relationship to science. The significance of the *purpose* of a scientific undertaking is, I believe, grossly underestimated. I would like to state a two-pronged thesis which in my estimation deserves consideration. Then I will elaborate the meaning of these two points.

1. In any scientific endeavor—whether "pure" or applied science—there is a prior personal subjective choice of the purpose or value which that scientific work is perceived as serving.

2. This subjective value choice which brings the scientific endeavor into being must always lie outside of that endeavor and can never become a part of the science involved in that endeavor.

Let me illustrate the first point from Dr. Skinner's writings. When he suggests that the task for the behavioral sciences is to make man "productive," "well-behaved," etc., it is obvious that he is making a choice. He might have chosen to make men submissive, dependent, and gregarious, for example. Yet by his own statement in another context man's "capacity to choose," his freedom to select his course and to initiate action—these powers do not exist in the scientific picture of man. Here is, I believe, the deepseated contradiction or paradox. Let me spell it out as clearly as I can.

Science, to be sure, rests on the assumption that behavior is caused—that a specified event is followed by a consequent event. Hence all is determined, nothing is free, choice is impossible. But we must recall that science itself and each specific scientific endeavor, each change of course in a scientific research, each interpretation of the meaning of a scientific finding, and each decision as to how the finding shall be applied rests upon a personal, subjective choice. Thus science in general exists in the same paradoxical situation as does Dr. Skinner. A personal, subjective choice made by man sets in motion the operations of science, which in time proclaims that there can be no such thing as a personal, subjective choice. I shall make some comments about this continuing paradox at a later point.

I stressed the fact that each of these choices, initiating or furthering the scientific venture, is a value choice. The scientist investigates this rather than that, because he feels the first investigation has more value for him. He chooses one method for his study rather than another because he values it more highly. He interprets his findings in one way rather than another because he believes the first way is closer to the truth, or more valid—in other words that it is closer to a criterion which he values. Now these value choices are never a part of the scientific venture itself. The value choices connected with a particular scientific enterprise always and necessarily lie outside of that enterprise.

I wish to make it clear that I am not saying that values cannot be included as a subject of science. It is not true that science deals only with certain classes of "facts" and that these classes do not include values. It is a bit more complex than that, as a simple illustration or two may make clear.

If I value knowledge of the "three R's" as a goal of education, the methods of science can give me increasingly accurate information as to how this goal may be achieved. If I value problem-solving ability as a goal of education, the scientific method can give me the same kind of help.

Now if I wish to determine whether problem-solving ability is "better" than knowledge of the three R's, then scientific method can also study those two values, but *only*—and this is very important—only in terms of some other value which I have subjectively chosen. I may value college success. Then I can determine whether problem-solving ability or knowledge of the three R's is more closely associated with that criterion. I may value personal integration or vocational success or responsible citizenship. I can determine whether problem-solving ability or knowledge of the three R's is "better" for achieving any one of these values. But the value or purpose which gives meaning to a particular scientific endeavor must always lie outside of that endeavor.

Though our concern here is largely with applied science, what I have been saying seems equally true of so-called pure science. In pure science the usual prior subjective value choice is the discovery of truth. But this is a subjective choice, and science can never say whether it is the best choice, save in the light of some other value. Geneticists in Russia, for example, had to make a subjective choice of whether it was better to pursue truth, or to discover facts which upheld a governmental dogma. Which choice is "better"? We could make a scientific investigation of those alternatives, but only in the light of some other subjectively chosen value. If, for example, we value the survival of a culture, then we could begin to investigate with the methods of science the question as to whether pursuit of truth or support of governmental dogma is most closely associated with cultural survival.

My point then is that any scientific endeavor, pure or applied, is carried on in the pursuit of a purpose or value which is subjectively chosen by persons. It is important that this choice be made explicit, since the particular value which is being sought can never be tested or evaluated, confirmed or denied, by the scientific endeavor to which it gives birth and meaning. The initial purpose or value always and necessarily lies outside the scope of the scientific effort which it sets in motion.

Perhaps, however, the thought is that a continuing scientific endeavor will evolve its own goals; that the initial findings will alter the directions, and subsequent findings will alter them still further; and that the science somehow develops its own purpose. This seems to be a view implicitly held by many scientists. It is surely a reasonable description, but it overlooks one element in this continuing development, which is that subjective, personal choice enters in at every point at which the direction changes. The findings of a science, the results of an experiment, do not and never can tell us what next scientific purpose to pursue. Even in the purest of science, the scientist must decide what the findings mean and must subjectively choose what next step will be most profitable in the pursuit of his purpose. And if we are speaking of the application of scientific knowledge, then it is distressingly clear that the increasing scientific knowledge of the structure of the atom carries with it no necessary choice as to the purpose to which this knowledge will be put. This is a subjective personal choice which must be made by many individuals.

Thus I return to the proposition with which I began this section of my remarks—and which I now repeat in different words. Science has its meaning as the objective pursuit of a purpose which has been subjectively chosen by a person or persons. This purpose or value can never be investigated by the particular scientific experiment or investigation to which it has given birth and meaning. Consequently, any discussion of the control of human beings by the behavioral sciences must first and most deeply concern itself with the subjectively chosen purposes which such an application of science is intended to implement.

## An Alternative Set of Values

If the line of reasoning I have been presenting is valid, then it opens new doors to us. If we frankly face the fact that science takes off from a subjectively chosen set of values, then we are free to select the values we wish to pursue. We are not limited to such stultifying goals as producing a controlled state of happiness, productivity, and the like. I would like to suggest a radically different alternative.

Suppose we start with a set of ends, values, purposes, quite different from the type of goals we have been considering. Suppose we do this quite openly, setting them forth as a possible value choice to be accepted or rejected. Suppose we select a set of values which focuses on fluid elements of process, rather than static attributes. We might then value:

Acc. No. 1899
Date 4. 8. 65.

Man as a process of becoming; as a process of achieving worth and dignity through the development of his potentialities;
The individual human being as a self-actualizing process, moving on to more challenging and enriching experiences;
The process by which the individual creatively adapts to an ever new and changing world;
The process by which knowledge transcends itself, as for example the theory of relativity transcended Newtonian physics, itself to be transcended in some future day by a new perception.

If we select values such as these, we turn to our science and technology of behavior with a very different set of questions. We will want to know such things as these.

Can science aid us in the discovery of new modes of richly rewarding living? More meaningful and satisfying modes of interpersonal relationships?

Can science inform us as to how the human race can become a more intelligent participant in its own evolution—its physical, psychological, and social evolution?

Can science inform us as to ways of releasing the creative capacity of individuals, which seem so necessary if we are to survive in this fantastically expanding atomic age? Dr. Oppenheimer has pointed out (8) that knowledge, which used to double in millennia or centuries, now doubles in a generation or a decade. It appears that we will need to discover the utmost in release of creativity if we are to be able to adapt effectively.

In short, can science discover the methods by which man can most readily become a continually developing and self-transcending process, in his behavior, his thinking, his knowledge? Can science predict and release an essentially "unpredictable" freedom?

154.4
1899
C.E.R.T., West Ben
Library

It is one of the virtues of science as a method that it is as able to advance and implement goals and purposes of this sort as it is to serve static values such as states of being well-informed, happy, obedient. Indeed we have some evidence of this.

### A Small Example

I will perhaps be forgiven if I document some of the possibilities along this line by turning to psychotherapy, the field I know best.

Psychotherapy as Meerloo (5) and others have pointed out can be one of the most subtle tools for the control of one person by another. The therapist can subtly mold individuals in imitation of himself. He can cause an individual to become a submissive and conforming being. When certain therapeutic principles are used in extreme fashion, we call it brainwashing, an instance of the disintegration of the personality and a reformulation of the person along lines desired by the controlling individual. So the principles of therapy can be used as a most effective means of external control of human personality and behavior. Can psychotherapy be anything else?

Here I find the developments going on in client-centered psychotherapy (*11*) an exciting hint of what a behavioral science can do in achieving the kinds of values I have stated. Quite aside from being a somewhat new orientation in psychotherapy, this development has important implications regarding the relation of a behavioral science to the control of human behavior. Let me describe our experience as it relates to the issues of today's discussion.

In client-centered therapy, we are deeply engaged in the prediction and influencing of behavior. As therapists we institute certain attitudinal conditions, and the client has relatively little voice in the establishment of these conditions. Very briefly, as I indicated earlier, we have found that the therapist is most effective if he is: (a) genuine, integrated, transparently real in the relationship; (b) acceptant of the client as a separate, different, person, and acceptant of each fluctuating aspect of the client as it comes to expression; and (c) sensitively empathic in his understanding, seeing the world through the client's eyes. Our research permits us to predict that if these attitudinal conditions are instituted or established, certain behavioral consequences will ensue. Putting it this way sounds as if we are again back in the familiar groove of being able to predict behavior, and hence able to control it. But precisely here exists a sharp difference.

The conditions we have chosen to establish predict such behavioral consequences as these: that the client will become more self-directing, less rigid, more open to the evidence of his senses, better organized and integrated, more similar to the ideal which he has chosen for himself. In other words we have established by external control conditions which we predict will be followed by internal control by the individual, in pursuit of internally chosen goals. We have set the conditions which predict various classes of behaviors—self-directing behaviors, sensitivity to realities within and without, flexible adaptiveness—which are by their very nature *unpredictable* in their specifics. The conditions we have established predict behavior which is essentially "free." Our recent research (*12*) indicates that our predictions are to a significant degree corroborated, and our commitment to the scientific method causes us to believe that more effective means of achieving these goals may be realized.

Research exists in other fields—industry, education, group dynamics—which seems to support our own findings. I believe it may be conservatively stated that scientific progress has been made in identifying those conditions in an interpersonal relationship which, if they exist in B, are followed in A by greater maturity in behavior, less dependence upon others, an increase in expressiveness as a person, an increase in variability, flexibility, and effectiveness of adaptation, an increase in self-responsibility and self-direction.

Thus we find ourselves in fundamental agreement with John Dewey's statement: "Science has made its way by releasing, not by suppressing, the elements of variation, of invention and innovation, of novel creation in individuals" (*10*). We have come to believe that progress in personal life and in group living is made in the same way, by releasing variation, freedom, creativity.

## A Possible Concept of the Control of Human Behavior

It is quite clear that the point of view I am expressing is in sharp contrast to the usual conception of the relationship of the behavioral sciences to the control of human behavior, previously mentioned. In order to make this contrast even more blunt, I will state this possibility in a form parallel to the steps which I described before.

1. It is possible for us to choose to value man as a self-actualizing process of becoming; to value creativity, and the process by which knowledge becomes self-transcending.

2. We can proceed, by the methods of science, to discover the conditions which necessarily precede these processes, and through continuing experimentation, to discover better means of achieving these purposes.

3. It is possible for individuals or groups to set these conditions, with a minimum of power or control. According to present knowledge, the only authority necessary is the authority to establish certain qualities of interpersonal relationship.

4. Exposed to these conditions, present knowledge suggests that individuals become more self-responsible, make progress in self-actualization, be-

come more flexible, more unique and varied, more creatively adaptive.

5. Thus such an initial choice would inaugurate the beginnings of a social system or subsystem in which values, knowledge, adaptive skills, and even the concept of science would be continually changing and self-transcending. The emphasis would be upon man as a process of becoming.

I believe it is clear that such a view as I have been describing does not lead to any definable Utopia. It would be impossible to predict its final outcome. It involves a step-by-step development, based upon a continuing subjective choice of purposes, which are implemented by the behavioral sciences. It is in the direction of the "open society," as that term has been defined by Popper (9), where individuals carry responsibility for personal decisions. It is at the opposite pole from his concept of the closed society, of which Walden Two would be an example.

I trust it is also evident that the whole emphasis is upon process, not upon end states of being. I am suggesting that it is by choosing to value certain qualitative elements of the process of becoming, that we can find a pathway toward the open society.

## The Choice

It is my hope that I have helped to clarify the range of choice which will lie before us and our children in regard to the behavioral sciences. We can choose to use our growing knowledge to enslave people in ways never dreamed of before, depersonalizing them, controlling them by means so carefully selected that they will perhaps never be aware of their loss of personhood. We can choose to utilize our scientific knowledge to make men necessarily happy, well-behaved, and productive, as Dr. Skinner suggests. We can, if we wish, choose to make men submissive, conforming, docile. Or at the other end of the spectrum of choice we can choose to use the behavioral sciences in ways which will free, not control; which will bring about constructive variability, not conformity; which will develop creativity, not contentment; which will facilitate each person in his self-directed process of becoming; which will aid individuals, groups, and even the concept of science to become self-transcending in freshly adaptive ways of meeting life and its problems.

If we choose to utilize our scientific knowledge to free men, then it will demand that we live openly and frankly with the great paradox of the behavioral sciences. We will recognize that behavior, when examined scientifically, is surely best understood as determined by prior causation. This is the great fact of science. But responsible personal choice, which is the most essential element in being a person, which is the core experience in psychotherapy, which exists prior to any scientific endeavor, is an equally prominent fact in our lives. That these two important elements of our experience appear to be in contradiction has perhaps the same significance as the contradiction between the wave theory and the corpuscular theory of light, both of which can be shown to be true, even though incompatible. We cannot profitably deny the freedom which exists in our subjective life, any more than we can deny the determinism which is evident in the objective description of that life. We will have to live with that paradox.

## REFERENCES

1. Asch, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
2. Beston, W. H., Heron, W., & Scott, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.*, 1954, **8**, 70–76.
3. Block, J., & Block, Jeanne. An investigation of the relationship between intolerance of ambiguity and ethnocentrism. *J. Pers.*, 1951, **19**, 303–311.
4. Crutchfield, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191–198.
5. Meerloo, J. A. M. Medication into submission: the danger of therapeutic coercion. *J. nerv. ment. Dis.*, 1955, **122**, 353–360.
6. Olds, J. A physiological study of reward. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955.
7. Oppenheimer, R. Analogy in science. *Amer. Psychologist*, 1956, **11**, 127–135.
8. Oppenheimer, R. Science and our times. [Chicago:] *Roosevelt University Occasional Papers*, 1956, 2.
9. Popper, K. R. *The open society and its enemies*. London: Rutledge and Kegan Paul, 1945.
10. Ratner, J. (Ed.) *Intelligence in the modern world: John Dewey's philosophy*. New York: Modern Library, 1939.
11. Rogers, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
12. Rogers, C. R., & Dymond, Rosalind F. (Eds.) *Psychotherapy and personality change*. Chicago: University of Chicago Press, 1954.
13. Seeman, J., & Raskin, N. J. Research perspectives in client-centered therapy. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research*, Chapter 9. New York: Ronald Press, 1953.
14. Skinner, B. F. Freedom and the control of men. *Amer. Scholar*, Winter, 1955–1956, **25**, 47–65.
15. Skinner, B. F. *Science and human behavior*. New York: Macmillan, 1953.
16. Skinner, B. F. *Walden Two*. New York: Macmillan, 1948.
17. Smith, G. J. W., Spence, D. F., & Klein, G. S. Subliminal effects of verbal stimuli. *J. abnorm. soc. Psychol.*, 1959, **59**, 167–176.
18. Taft, R. Intolerance of ambiguity and ethnocentrism. *J. consult. Psychol.*, 1956, **20**, 153–154.
19. Thetford, W. N. An objective measure of frustration tolerance in evaluating psychotherapy. In W. Wolff (Ed.), *Success in psychotherapy*, Chapter 2. New York: Grune and Stratton, 1952.

# 2

# What Psychology Can We Trust?

## — Goodwin Watson

*The answer to Watson's question, "What psychology can we trust?" might well be, "All of them!" Drawing from the varied approaches and methods used by different "schools" in the study of behavior, the author attempts to demonstrate the wide range of applicability in the classroom of psychological principles and theoretical explanations. The role of science in developing means to reach goals is especially evident here.*

EDUCATORS AND others who wish to apply psychology in their professional work have long been troubled by controversies among psychologists themselves. Behaviorism arose to challenge the introspective method; Thorndike's connectionism was controverted by Gestalt concepts; psychoanalysts talked an almost completely different language. It was natural for teachers to say, "Let's wait until the psychologists themselves straighten out their various systems!" It looked for a while as if one could support almost any educational practice by citing the appropriate psychologist.

Gradually, however, a body of fairly firm facts has accumulated. While it remains true that research findings will be somewhat differently expressed and explained within different theoretical frameworks, the findings themselves are fairly solid.

---

Reprinted from the booklet *What Psychology Can We Trust?* Bureau of Publications, Teachers College, Columbia University, 1961, by permission of the publisher. This booklet is a revision and expansion, by the author, of his article entitled "What Psychology Can We Feel Sure About?" published in the *Teachers College Record*, 1960, **61** (5).

A workshop of educators recently asked me to formulate for them some statements of what we really know today about children and learning. I set out to list some propositions, important for education, with which I thought few knowledgeable psychologists of any "school" would disagree. To my own surprise, the list grew to fifty.

In no science are truths established beyond the possibility of revision. Einstein modified thinking about gravity, even though Newton's observations were essentially correct. Psychology is much younger and more malleable than physics. New facts are constantly accumulating in psychological research, and these will doubtless introduce some qualifications and modifications—conceivably even a basic contradiction. The educator who bases his program on the propositions presented here, however, is entitled to feel that he is on solid psychological ground and not on shifting sands. Most psychologists, I think, would agree with them.

The fifty propositions follow, somewhat revised and with brief comments, under nine major headings.

## I. Learning Process

*1. Behaviors which are rewarded (reinforced) are more likely to recur.* This most fundamental law of learning has been demonstrated in literally thousands of experiments. It seems to hold for every sort of animal from earthworms to highly intelligent adults. The behavior most likely to emerge in any situation is that which the subject found successful or satisfying previously in a similar situation. No other variable affects learning so powerfully. The best-planned learning provides for a steady, cumulative sequence of successful behaviors.

*2. Reward (reinforcement), to be most effective in learning, must follow almost immediately after the desired behavior and be clearly connected with that behavior in the mind of the learner.* The simple word, "Right," coming directly after a given response, will have more influence on learning than any big reward which comes much later or which is dimly connected with many responses so that it can't really reinforce any of them. Much of the effectiveness of programmed self-instruction (teaching machines) is that success is fed back immediately for each learner response. A total mark on a test, the day after it is administered, has little or no reinforcement value for the specific answers.

*3. Sheer repetition without indications of improvement or any kind of reinforcement (reward) is a poor way to attempt to learn.* Practice is not enough. The learner cannot improve by repeated efforts unless he is informed about how well each one has succeeded.

*4. Threat and punishment have variable and uncertain effects upon learning; they may make the punished response more likely or less likely to recur; they may set up avoidance tendencies which prevent further learning.* Punishment is not, psychologically, the reverse of reward. It disturbs the relationship of the learner to the situation and teacher. It does not assist the learner in finding and fixing the correct response.

### II. Motivation

*5. Readiness for any new learning is a complex product of interaction among such factors as (a) sufficient physiological and psychological maturity, (b) sense of the importance of the new learning for the learner in his world, (c) mastery of prerequisites providing a fair chance of success, and (d) freedom from discouragement (expectation of failure) or threat (sense of danger).* Conversely, the learner will not be ready to try new responses which are beyond his powers or which are seen as valueless or too dangerous.

*6. Opportunity for fresh, novel, stimulating experience is a kind of reward which is quite effective in conditioning and learning.* Experiments indicate that even lower animals (rats, dogs, monkeys) will learn as effectively when they receive rewards of new experience or satisfied curiosity, as they will when the rewards gratify desires for food, water, sex, or rest. One trouble with the typical school day is that it is too dull; pupils find too few stimulating new insights to reward their efforts.

*7. The type of reward (reinforcement) which has the greatest transfer value to other life-situations is the kind one gives oneself—the sense of satisfaction in achieving purposes.* Any extrinisic reward—candy, or stars on a chart, or commendation—depends on its dispenser. There is no need to strive if the reward-giver is out of the picture. Also, cheating (detour behavior) can sometimes win the extrinsic reward. The internal reward system is always present for the learner, and he sees little gain in fooling himself.

*8. Learners progress in any area of learning only as far as they need to in order to achieve their purposes. Often they do only well enough to "get by"; with increased motivation they improve.* One of the earliest psychological studies showed that telegraph operators with years of practice seldom rose in speed beyond the acceptable "main line" rate; they could, however, learn to send and receive much faster if special incentives were provided. Studies of reading speed show that practice alone will not bring improvement; a person may have read books for years at his customary rate, but with new demands and opportunities he may be able to double that rate.

*9. The most effective effort is put forth by children when they attempt* *tasks which fall in the "range of challenge"—not too easy and not too hard—where success seems quite possible but not certain.* A good illustration is the dart game. Where is it most fun to stand? So close that you can't miss? So far away that you can't make a decent score? The maximum fun is at the level of maximum challenge. If it proves too easy, one steps further back; if it proves too hard one steps a little closer. It is not reasonable to expect a teacher to set an appropriate level of challenge for each pupil in a class; pupils can, however, learn to set their own goals to bring maximum satisfaction and learning.

## III. Teaching Methods

10. *Children are more apt to throw themselves wholeheartedly into any project if they themselves have participated in the selection and planning of the enterprise.* Genuine participation (not pretended sharing) has been found to increase motivation, adaptability, and speed of learning.

11. *Reaction to excessive direction by the teacher is likely to be (a) apathetic conformity, (b) defiance, (c) scape-goating, or (d) escape from the whole affair.* Autocratic leadership has been found to increase dependence of members on the leader and to generate resentment (conscious or unconscious) which finds expression in attacks on weaker figures or even in sabotage of the work.

12. *Over-strict discipline is associated with more conformity, anxiety, shyness, and acquiescence in children; greater permissiveness is associated with more initiative and creativity in children.* Comparison of children whose parents were at the most permissive extreme of home discipline with those who were most strict (both groups of parents loving and concerned) has shown more enterprise, self-confidence, curiosity, and originality in the youngsters from permissive homes. These children were also more friendly and less hostile than the strictly disciplined children. The demand for more restrictive discipline runs counter to the call for more individuality.

13. *Many pupils experience so much criticism, failure, and discouragement in school that their self-confidence, level of aspiration, and sense of worth are damaged.* The pupil who sees himself at his worst in school is likely to place little value on study and to seek his role of importance outside the classroom. He may carry through life a sense of being not good for much. He is likely also to feel resentment at schools, teachers, and books.

14. *When children (or adults) experience too much frustration, their behavior ceases to be integrated, purposeful, and rational. Blindly they act out their rage or discouragement or withdrawal. The threshold of what is "too much" varies; it is lowered by previous failures.* Pupils who have had little

success and almost continuous failure at school tasks are in no condition to think, to learn, or even to pay attention. They may turn their anger outward against respectable society or inward against themselves.

15. *Pupils* think *when they encounter an obstacle, difficulty, puzzle, or challenge in a course of action which interests them. The process of thinking involves designing and testing plausible solutions for the problem as understood by the thinker.* It is useless to command people to think; they must feel concerned to get somewhere and eager to remove an obstruction on the way. A situation in which the customary responses work smoothly does not call for thought; when they no longer serve, then one needs to think.

16. *The best way to help pupils form a general concept is to present the concept in numerous and varied specific situations, contrasting experiences with and without the desired concept, then to encourage precise formulations of the general idea and its application in situations different from those in which the concept was learned.* For example, the concept of democracy might be illustrated not only in national government but also in familiar situations of home, school, church, jobs, clubs, and local affairs. It is best understood when it is contrasted with other power structures such as autocracy, oligarchy, or laissez faire.

17. *The experience of learning by sudden insight into a previously confused or puzzling situation arises when (a) there has been a sufficient background and preparation, (b) attention is given to the relationships operative in the whole situation, (c) the perceptual structure "frees" the key elements to be shifted into new patterns, (d) the task is meaningful and within the range of ability of the subject.* The term "cognitive reorganization" is sometimes applied to this experience. Suddenly the scene changes into one that seems familiar and can be coped with.

18. *Learning from reading is facilitated more by time spent recalling what has been read than by rereading.* In one experiment (typical of many), students who spent 80 per cent of their learning periods trying to remember what they had read surpassed those who spent only 60 per cent of the time on recollection; 60 per cent was better than 40 per cent; 40 per cent was better than 20 per cent; and the poorest plan was to spend all the time reading and rereading the assignment. Trying to interpret and apply the ideas noted in reading is another good procedure.

19. *Forgetting proceeds rapidly at first—then more and more slowly; recall shortly after learning reduces the amount forgotten.* Within twenty-four hours after learning something, a large part is forgotten unless efforts are made to prevent forgetting. A thing can be relearned more quickly than it was learned originally, however; and if it is reviewed several times at gradually increasing intervals it can be retained for some time. Some people apply this idea in remembering names of several persons to whom they have just

been introduced. The first recall must come immediately; a second recall should follow about five minutes later, a third after half an hour, a fourth some hours later, and perhaps a fifth the next day. Then names will stick.

20. *People remember new information which confirms their previous attitudes better than they remember new information which runs counter to their previous attitudes.* Studies consistently show that individuals who feel strongly on a controversial issue, and who are asked to read presentations of both sides, remember the facts and arguments which support their feelings better than they recall those on the opposite side.

## IV. Subject Matter

21. *No school subjects are markedly superior to others for "strengthening mental powers." General improvement as a result of study of any subject depends on instruction designed to build up generalizations about principles, concept formation, and improvements of techniques of study, thinking, and communication.* In Thorndike's classic investigation, high school pupils whose programs included "hard" subjects like mathematics and ancient languages did not consistently improve in "selective and relational thinking" during a school year any more than did students whose programs had substituted allegedly easier activities like dramatics or typewriting. Whether any subject develops general abilities or not depends very much on how it is taught. Memorizing won't strengthen memory, but students can learn how to use better techniques of remembering. Pupils who study science may or may not master the processes of scientific thought.

22. *Children (and adults even more) tend to select groups, reading matter, TV shows, and other influences which agree with their own opinions; they break off contact with contradictory views.* Parents want children taught what they themselves value and believe. One of the basic educational problems is to preserve minds from being closed in—surrounded by like-minded associates, like-minded commentators, and like-minded publications.

23. *What is learned is most likely to be available for use if it is learned in a situation much like that in which it is to be used and immediately preceding the time when it is needed. Learning in childhood, then forgetting, and then relearning when need arises is not an efficient procedure.* It was once thought that childhood was the golden age for learning. We now know (see Proposition 33) that adults of forty can learn better than youths of fourteen and much better than seven-year-olds. The best time to learn is when the learning can be useful. Motivation is then strongest and forgetting less of a problem. Much that is now taught children might be more effective if taught to responsible adults.

*24. The superiority of man over calculating machines is more evident in the formulation of questions than in the working out of answers.* Is education too much concerned with answers, not enough with the generation of better questions? "Always the beautiful answer who asks a more beautiful question" (E. E. Cummings).

*25. Television is the most frequently reported activity of elementary school pupils, occupying about the same number of hours per week as are given to school—far more than would voluntarily be given to school attendance.* Television is qualitatively as well as quantitatively influential. The pictures and sound have a lifelike impact, far more impressive than print. The proportion of their total knowledge which children today acquire in school has dropped to a small fraction. Communities deeply concerned about finding better textbooks have given very little effective attention to the more potent educational medium.

## V. Evaluation

*26. If there is a discrepancy between the real* objectives *and the* tests *used to measure achievement, the latter become the main influence upon choice of subject matter and method.* The biggest difficulty with our present standard tests, and with the machines proposed for offering programmed instruction, is that they don't cover what we believe to be our most important educational aims and values. Curriculum and teaching geared to such tests and machines are likely, then, to ignore these great goals, and to concentrate only on learnings which can be easily checked and scored.

## VI. Growth

*27. Every trait in human behavior is a product of the interaction of heredity (as determined at conception by genes) and environmental influences.* The classic argument on the relative importance of heredity and environment is futile; both influence every human act and trait. Some traits (preferences in food or clothing, for example) are easily influenced by nurture, others (height, rate of skeletal ossification) only when extreme environmental differences are involved.

*28. There are specific stages in individual development during which certain capacities for behavior appear. The manner in which these capacities are then utilized sets a pattern for later behavior which is highly resistant to change. If unutilized then, they are likely not to develop later.* The object which a baby duck regards as its mother and follows about is that which appears at a certain brief stage shortly after the egg has hatched. Children born blind who later recover their sight have lifelong difficulty recognizing objects which seem simple and obvious to youngsters with normal visual development. A child who does not experience affectionate care from one person (not necessarily the biological mother) during the period from six months to eighteen months of age may be emotionally handicapped for life. It is very difficult in later years to learn to speak a second language without a foreign accent. Adolescents who do not develop independence from parents and a normal interest in the opposite sex during the 'teen years may never do either very well.

29. *The most rapid mental growth occurs during infancy and early childhood; the average child achieves about half of his total mental growth by the age of five.* In the first two years a normal child transforms the "big, buzzing, blooming confusion" of his first conscious experience to organized perception of familiar faces, spoken words, surroundings, toys, bed, clothing, and foods. He differentiates himself from others, high from low, many from few, approval from disapproval. He lays a foundation for lifelong tendencies toward trust or mistrust, self-acceptance or shame, initiative or passivity; and these vitally condition further mental growth.

30. *During the elementary school years (ages six to twelve) most children enjoy energetic activity—running, chasing, jumping, shouting, and roughhouse. For most staid adults this is uncomfortable. Boys are generally more vigorous, active, rough, noisy, and non-conforming than are girls.* Only in relatively recent periods of human history have children been cooped up in classrooms, confined to desks, and marched in silent order. There is reason to doubt the biological appropriateness of the traditional school.

31. *Not until adolescence do most children develop the sense of time which is required for historical perspective.* The so-called facts of history—1492, 1776, and all that—can be learned by children but without any real grasp of what life was like in another period or in a different country. Most instruction in Biblical, ancient, medieval, and even modern history is no more real to children than are fairy tales.

32. *The significance of the important biological transformations of pubescence (growth of primary sex organs, development of secondary sex characteristics, skeletal and muscular growth, glandular interaction) lies mainly in the* meaning *which cultural norms and personal history have given to these changes.* Adolescence is not necessarily a "stormy" period. It is vital in our culture because it is then that we expect boys and girls to achieve their *identity.* They must work out their vocational plans, their social life, their marriage, their politics, their religion, and, above all, their concept of who and what they really are. They have one role at home, another with friends of their own sex, another with friends of the opposite sex, another at

work, and still another in a school class. What is the true self behind all these masks?

33. *Ability to learn increases with age up to adult years.* The apparent decline is more a matter of motivation (see Proposition 8) than of inability. Most adults don't need to take the trouble to change and to learn new skills. We can coerce children into school activities; adult education is mostly voluntary. Men and women *can*, if they wish, master new languages, new ideas, and new ways of acting or problem-solving even at sixty and seventy years of age.

34. *No two people make the same response to any school situation. Differences of heredity, physical maturity, intelligence, motor skills, health, experiences with parents, siblings, playmates; consequent attitudes, motives, drives, tastes, fears—all these and more enter into production of each individual's unique reaction. People vary in their minds and personalities as much as in their appearance.* Because conformity can be coerced, the ever-present differences are often ignored. Almost everyone around us stops at the red light, comes to work on time, sees the favorite TV programs. Children assemble in classes and take out their arithmetic books when told to do so. These similarities are misleading; the inner reactions are always varied.

35. *Pupils vary not only in their present performance but in their rate of growth and the "ceiling" which represents their potential level of achievement. Some "late bloomers" may eventually surpass pupils who seem far ahead of them in grade school.* Rates of growth are, moreover, irregular; mental growth is even more so than physical growth because it varies with encouragement and opportunity.

36. *Children's gains in intelligence test scores and IQ's are positively related to aggressiveness, competitiveness, initiative, and strength of felt need to achieve.* Some apparent dullness is due to lack of encouragement in the development of verbal and arithmetic skills. When these are not valued at home or by friends, it is hard to take them very seriously. Studies have shown gradual decline in IQ's in culturally deprived groups.

37. *Pupils grouped by ability on any one kind of test (age, size, IQ, reading, arithmetic, science, art, music, physical fitness, and so forth) will vary over a range of several grades in other abilities and traits.* Homogeneous grouping is literally impossible except for a given limited task. On another task, a little later, these same children will not perform alike. No teacher is justified in teaching children, however they may be grouped, as if they were all alike—little mechanical robots. A challenging assignment (see Proposition 9) for one pupil may be too hard or too easy for another despite similarity in grade, age, IQ, sex, size, and school marks.

## VIII. Group Relations

38. *The right size of group for any activity depends on both the maturity of the individuals and the nature of the activity.* Hundreds or thousands may be spectators at a film, a TV presentation, or a spectacle. Working, interacting groups seem to do best when composed of five to eight members. If the group is larger, some become performers and other spectators. At age six, spontaneous groups seldom exceed three or four children. Sizes now accepted for school classes are much too large for good cooperative work.

39. *When groups act for a common goal there is better cooperation and more friendliness than when individuals in the group are engaged in competitive rivalry with one another.* Some studies indicate that the more cooperative groups also produce results of better quality. The competitive emphasis directs attention toward winning rather than toward excellence of performance.

40. *Children learn at an early age that peer consensus is an important criterion; they are uncomfortable when they disagree with their peers, and especially when they find themselves in a minority of one against all the others.* Children naturally conform; they do not want to be seen as "different" from all the others. This is an important aid in socializing the human animal. "Independence" emerges in some adolescents as conformity to a specially selected reference group—often one not present at the moment.

41. *Pupils learn much from one another; those who have been together for years learn new material more easily from one of their own group than they do from strangers.* Traditioned groups develop their own customary norms. They have signals which recall shared experiences. Peers can communicate economically and without barriers of status.

42. *Children who differ in race, nationality, religion, or social class background, but who play together on a footing of equal status and acceptance, usually come to like one another.* This is a particular instance of an important law in social psychology: frequency of interaction brings increased liking for one another. Generally, best friends are those who are most frequently seen—classmates, neighbors, work associates, companions in recreation. The relationship is circular. Those best liked are most often chosen as companions; those with whom we more frequently associate become better liked.

43. *In most school classes, one to three pupils remain unchosen by their classmates for friendship, for parties, or for working committees. These "isolates" are usually also unpopular with teachers.* A vicious circle operates. The rejection by peers and teachers intensifies the disagreeable characteristics and behavior of the isolates and makes it even more difficult to accept

them. The task of the teacher is to break through the vicious circle—to try to bring out traits which will enable the isolate eventually to make friends. The relationship to an accepting and understanding teacher is the bridge to all other human beings.

44. *Leadership qualities vary with the demands of the particular situation. A good leader for a football team may or may not be a good leader for a discussion group, a research project, or an overnight hike; leadership is not a general trait.* In some groups, leadership is distributed among members. One person may propose an idea; someone else may lead in its development; a third person may remind the group of its goals or time limits; still another may help to raise morale. No one leader is required. Agenda may be adopted by consensus and followed by cooperative reminders as needed.

## IX. Social Stratification

45. *The socioeconomic class into which a child happens to be born strongly influences his life chances.* Studies show differential infant-mortality rates, health services, mental growth, school performance, housing, recreation, vocational opportunities, type of church, and extent of community influence—all favoring the upper-class neighborhoods. Schools have a peculiar problem with social class: the school boards represent largely the upper strata; the teachers are drawn mainly from the middle layers; most pupils come from the lower classes.

46. *Two thirds of the elementary school children of America come from lower-class homes; the one third who come from the lower-lower class usually find school very uncongenial.* This is education's biggest unsolved problem. Even in the elementary schools, most teachers are reluctant to work in the culturally deprived areas of cities or of rural regions. As we try to keep all American adolescents in school until sixteen or eighteen years of age, we increase the problem. The "Blackboard Jungle" of the city slums and the "Tobacco Road" schools of the rural slums are tough for teachers and are unrewarding for the resistant pupils. Few successful experiments have yet been reported in devising publicly approved educational programs with strong appeal for these students and for teachers.

47. *Children who are looked down upon (or looked up to) because of their family, school marks, social class, race, nationality, religion, or sex tend to adopt and to internalize this evaluation of themselves.* Studies have shown that Negro children in a white-dominated society are more likely to pick the black-skinned doll than the white-skinned doll as "looking bad"; they unconsciously accept the prevailing norms of white superiority. Race

and class and sex discrimination lead to self-concepts of over-valuation in the dominant group and under-valuation in the subordinate group.

48. *Attitudes toward members of out-groups are usually acquired from members of one's in-group.* White children get their attitudes toward Negroes not primarily from direct contact with Negroes but from their white parents and playmates. Similarly, attitudes toward other classes, other faiths, and other nationalities are acquired (like speech dialects or food preferences) from the nurturing group.

49. *Children choose most of their "best friends" from homes of the same socioeconomic class as their own.* An effort in one junior high school over a three-year period to increase interaction across the lines of socioeconomic class led to a drop in same-class choices from 75 to 66 per cent. The norms of the larger community were more powerful than those generated within the school.

50. *More girls than boys wish, from time to time, that they could change their sex.* Sex discrimination in our culture has received much less attention than has race discrimination. While there has been some progress toward equality of opportunity in education, our culture still sees boys as primarily concerned with vocational achievement and only secondarily with love, marriage, and parenthood. For girls, the order is reversed; it is thought that their vocational achievement should properly be secondary. Thus it comes about that the top positions in government, industry, commerce, science, education, and religion are reserved for males. The conflict may be particularly acute in girls gifted with high intelligence and outstanding capacity for executive leadership.

# 3

# The Relevancy of Educational Psychology

## — *Arthur P. Coladarci*

*The author conceives of the teacher's role as helping the learner change his behavior in specified desirable directions. Given this conception, teaching involves considerably more than a cookbook-like application of principles. It requires a creative and dynamic inquiry into teaching methods and procedures—a creative inquiry because every student, classroom, and school presents a unique situation; dynamic because the teacher must continuously inquire into the validity of his own procedures and be ready to modify them. He must formulate hypotheses about the most effective way to go about attaining goals. The contributions of educational psychology are held to be relevant in two ways: in the generation of educational hypotheses and in the acquisition of attitudes and skills necessary to intelligent hypothesizing and to the testing of hypotheses.*

THE RELEVANCY of an applied area depends in part upon the definition of the process, institution, or event to which it is applied. The contribution that can be made by *educational* psychology is partially a function of the particular meaning invested in "education." This statement is not merely the usual innocuous preface to an extended discussion. Indeed, it is our major thesis. Too many teachers and administrators have thought of educational psychology as consisting only of an ordered catalogue of educational prescriptions, which, together with those provided by the other foundational

Reprinted from *Educational Leadership*, May, 1956, 13 (8), 489–492, by permission of the author and the Association for Supervision and Curriculum Development.

fields in education, "tell" the teacher "how to teach" and the administrator "how to administer." The fallacy lies not only in the much too complimentary respect for the status of our knowledge in these areas but, more fundamentally, in the conception of education as a collection of successful recipes—the teacher or administrator is a person who has been armed with a bag-of-tricks into which he reaches for a decision regarding any given specific professional problem. Although this unfortunate orientation becomes an increasingly less frequent one, it still exists and may be partially attributable to the turn-of-the-century efforts to make education "scientific" by attempting to make it merely more *factual* (1).

If one, however, thinks of the nature of the educator's role in another way, educational psychology, and education generally, become more powerful, exciting, and rigorous. The conception we have in mind can be described by beginning with a rather coarse but generally acceptable definition of the educator's role: to help the learner change his behavior in specified desirable directions. Although the definition is too ambiguous for detailed analysis, it serves to point out the two basic factors involved: a *process* ("behavior change") and a *criterion* ("specified desirable directions"). Suppose that the educator has clearly specified what he means by "desirable" behavior changes in the form of operationally stated educational goals (2). It appears, now, that the focal task for the teacher is to so interact with his pupils, and to so arrange the conditions and materials, that these pupils will change in the hoped-for ways. Put in these terms, the teacher's task can be seen as one of manipulating the learning situation in such a way that the *predicted* behavior changes actually do occur. If, at this point, the educational psychologist could say that we now know which manipulations will produce the desired changes, no problem would exist—we have only to apply the correct recipe. However, educational psychology cannot do this. Any particular combination of teacher-pupil-class-group-community-available materials, etc., is somewhat different from any other combination. There is no general prescription that can be considered to be clearly valid for particular cases. The teacher, then, *must be an active, continuous inquirer into the validity of his own procedures.* As Corey puts it:

> Most of the study of what should be kept in the schools and what should go and what should be added must be done in hundreds of thousands of classrooms and thousands of American communities. The studies must be understood by those who may have to change the way they do things as a result of the studies. Our schools cannot keep up with the life they are supposed to sustain and improve unless teachers, pupils, supervisors, administrators, and school patrons continuously examine what they are doing. Singly and in groups, they must use their imagination creatively and constructively to identify the practices that must be changed to meet the needs and demands of modern life, courageously to try out those prac-

tices that give better promise, and methodically and systematically gather evidence to test their worth (3).

At the risk of belaboring the point, let us put it in somewhat different form before considering the relevancy of educational psychology. The educator's decisions about methods, materials, and curricular procedures should be thought of as *hypotheses* regarding the way in which the desired behavior changes can be brought about. These hypotheses must be *tested* continuously by inquiring into the degree to which the predicted behavior changes actually occurred. This view has been referred to elsewhere by the writer (4) as "teaching behavior defined as the testing-of-hypotheses behavior." The crucial element is *tentativeness;* ideas and decisions about method and curriculum are to be held hypothetically, continuously tested, and continuously revised if necessary.

## Contribution of Educational Psychology

Given this conception of the educator's role, how can educational psychology be brought to bear on it in helpful ways? The contribution can be broken down into two related categories. First, educational psychology, as a body of information and an arena of research activity, can help in the generation of the educational hypotheses. Intelligent hypotheses are not chosen randomly nor are they found full-blown. An intelligent hypothesizer thinks along the lines of the following model: "*On the basis of the best information now available to me, I* hypothesize that this procedure will produce this result." To translate this into the context of education, we might say, for instance: "*On the basis of what* I *now know* about individual differences and the reading process, I hypothesize that this kind of grouping-for-reading will lead to the kind of pupil progress in reading that I would like to bring about."

Educational psychology, as a source of information, contributes to the "on-the-basis-of-what-I-now-know" portion of the statement. It helps provide information on which to base hypotheses for particular purposes and particular children. The teacher or administrator who takes this point seriously will understand that one cannot merely "take a course in educational psychology," but that he must constantly keep informed about those developments in this area that are most relevant to his particular educational responsibilities. The reader may also note that this conception of the interaction between educational psychology and the teacher means that every teacher can *contribute* to educational psychology in the process of testing his hypotheses.

A second kind of contribution which educational psychology can make is that of helping teachers and administrators to acquire the attitudes and skills necessary to intelligent hypothesizing and the testing of hypotheses. Limitations of space preclude an explication of this. Generally, what is involved is learning such skills as how to interpret data intelligently, how to observe accurately, how to avoid common logical fallacies in making inferences, how to make adequate decisions regarding what data should be gathered, ways in which data can be gathered and recorded, etc.

Both of these contributions of educational psychology are shared by all the fields represented in this symposium. In the writer's view, this is the *raison d'être* of any field that purports to be "foundational" in professional education. Educational psychology, of course, has many additional and somewhat unique values for the educator. We have chosen to overlook those in this discussion since they are covered comprehensively and in detail in the available published literature. Those who are interested are invited to examine the published reports of a committee organized by the Executive Committee of the National Society of College Teachers of Education. The first report (5) discussed the ways in which educational psychology relates to curriculum development; the second (6) considers the nature of educational psychology and its general place in teacher education; the third (7) gives detailed attention to the ways in which specific areas of educational psychology can be helpful to the prospective teacher; the last report (8) describes present practices and developments in the teaching of educational psychology.

It is appropriate, in this case, that the final comment should be cautionary as well as benedictory. The writer has stated his position as though there are no responsible competing alternatives to it. Any dogmatic flavor in the statement is more a consequence of brevity than of intent. Many persons will hold that such a conception of education as we have presented here is both impractical and not valuable. Our response would be that the orientation is at least practical in the sense that many, many educators have learned to behave as inquirers; the orientation appears to be valuable in that where one finds such an educator, he usually finds him to be valued by his colleagues, ego-involved in his profession, and able to criticize his procedures rationally. In short, such educators do exist and they appear to make the profession a better one by their membership in it.

## REFERENCES

1. Smith, B. O. Science of education. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1950. Pp. 1145–1152.

2. Travers, R. M. W. *Educational measurement.* New York: Macmillan, 1955. Pp. 19–36.
3. Corey, S. M. *Action research to improve school practices.* New York: Bureau of Publications, Teachers College, Columbia Univer., 1953. P. viii.
4. Coladarci, A. P. Are Educational researchers prepared to do meaningful research. *Calif. J. educ. Res.*, 1954, 5, 3–6.
5. The psychological basis of the modern curriculum. *J. educ. Psychol.*, 1948, **39,** 129–169.
6. Educational psychology in the education of teachers. *J. educ. Psychol.*, 1949, **40,** 257–294.
7. Educational psychology for teachers. *J. educ. Psychol.*, 1950, **41,** 321–372.
8. Current practices and innovations in the teaching of educational psychology. *J. educ. Psychol.*, 1952, 43, 1–30.

# Part II

## *Learner Abilities and Characteristics*

*ONE IMPORTANT set of factors to be considered in planning experiences intended to facilitate efficiency of student learning concerns the nature of the abilities and characteristics of the learners themselves. Awareness of the differences and similarities among learners of different sexes, ages, etc., will enable us to plan more effective learning experiences. In Part II: Learner Abilities and Characteristics, eight selections related to this set of factors are presented.*

*In the first selection, J. P. Guilford, professor of psychology at the University of Southern California, discusses the range of human intellectual abilities under the rubric "Structure of intellect." The major implication of his theoretical model is that to know an individual's intellectual resources thoroughly, we need a surprisingly large number of test scores.*

*In the next selection, reading number five, Katherine Bradway, of Stanford University, and Clare Thompson, of Washington State University, report a longitudinal study of intelligence. Though not based on Guilford's model, their study supports the need for a multiscore approach to the assessment of intelligence. Also, their results challenge earlier conclusions that mental growth levels off during early adolescence and remains constant thereafter.*

*The relationship of anxiety both to IQ and to achievement is investigated by Herbert Klausmeier and John Feldhusen in the sixth reading. The authors, professors of educational psychology at the Universities of Wisconsin and Purdue, respectively, used elementary-school children as subjects. Their research exemplifies the importance of obtaining information about the relationship between affective characteristics and cognitive abilities in learners.*

*Leonore Boehm, of Brooklyn College, explores the nature of children's thinking in the next selection. Emphasis is on differences in thinking characteristics at different developmental levels. This discussion, based on Piaget's investigations of children's thinking, has important implications for curriculum planning and for methods of teaching.*

*Reading number eight, "Education and the Psychology of Individual Differences," is by Philip E. Vernon, professor of educational psychology at the University of London's Institute of Education and editor of the British Journal of Educational Psychology. In the article Vernon reviews and analyzes the many practices and proposals, current in both England and the United States, designed to provide for differences in intellectual ability among students. The article was written while the author was a visting lecturer at Princeton University and a research fellow at the Educational Testing Service.*

*The effects of acceleration, as one method designed to provide for differences in intellectual ability among students, is investigated in the ninth reading by Herbert Klausmeier and Richard Ripple, who is professor of educational psychology at Cornell University. Results from this study, based on the performances of subjects on 32 measures, indicated no unfavorable academic, social, emotional, or physical correlates when second-grade pupils of superior learning abilities were accelerated to the fourth grade.*

*Shifting to characteristics of learners in the affective domain, the next selection reports an investigation by Dale Harris of the life problems and interests of adolescents in 1935 and 1957. Harris, professor of psychology at Pennsylvania State University, attempts to uncover changes in the relative emphasis given to separate problems and interests by adolescents as a result of changes in the social and economic structures of society during this 22-year period.*

*The final selection of Part II, entitled "The Adolescent Subculture and Academic Achievement," is by James Coleman, of Johns Hopkins University. Coleman discusses the sources of the adolescent's value system, especially the effects of interscholastic athletics on adolescent values. His suggestions regarding interscholastic academic competition—that to encourage it would help in effect to increase the value of academic achievement—are based on inferences drawn from the effects of interscholastic athletics.*

# 4

# Three Faces of Intellect — *J. P. Guilford*

*Classroom teachers and other professionals involved in the educational enterprise are constantly dealing with human intellectual abilities. Often they are called upon to interpret and apply the results of various devices designed to measure "intelligence." The IQ score has become a popular term that refers to the whole of an individual's intellectual resources. The author attempts to analyze human intelligence into its components by presenting a unified theory of human intellect. The theory is presented in the form of a model, called the "structure of intellect," that organizes the known, unique, or primary intellectual abilities into a single system. Examples of tests purporting to measure the various abilities represented in the model are given. The dangers in relying exclusively on a single index of intelligence should be apparent after reading this article. The establishment of a multiplicity of intellectual abilities has provocative implications for psychological theory, vocational testing, and education.*

MY SUBJECT is in the area of human intelligence, in connection with which the names of Terman and Stanford have become known the world over. The Stanford Revision of the Binet intelligence scale has been the standard against which all other instruments for the measurement of intelligence have been compared. The term IQ or intelligence quotient has become a household word in this country. This is illustrated by two brief stories.

A few years ago, one of my neighbors came home from a PTA meeting,

Reprinted from *American Psychologist*, 1959, **14**, 469–479, by permission of the author and the American Psychological Association.

remarking: "That Mrs. So-And-So, thinks she knows so much. She kept talking about the 'intelligence *quota*' of the children; 'intelligence *quota*'; imagine. Why, everybody knows that IQ stands for 'intelligence *quiz*.'"

The other story comes from a little comic strip in a Los Angeles morning newspaper, called "Junior Grade." In the first picture a little boy meets a little girl, both apparently about the first-grade level. The little girl remarks, "I have a high IQ." The little boy, puzzled, said, "You have a what?" The little girl repeated, "I have a high IQ," then went on her way. The little boy, looking thoughtful, said, "And she looks like such a nice little girl, too."

It is my purpose to speak about the analysis of this thing called human intelligence into its components. I do not believe that either Binet or Terman, if they were still with us, would object to the idea of a searching and detailed study of intelligence, aimed toward a better understanding of its nature. Preceding the development of his intelligence scale, Binet had done much research on different kinds of thinking activities and apparently recognized that intelligence has a number of aspects. It is to the lasting credit of both Binet and Terman that they introduced such a variety of tasks into their intelligence scales.

Two related events of very recent history make it imperative that we learn all we can regarding the nature of intelligence. I am referring to the advent of the artificial satellites and planets and to the crisis in education that has arisen in part as a consequence. The preservation of our way of life and our future security depend upon our most important national resources: our intellectual abilities and, more particularly, our creative abilities. It is time, then, that we learn all we can about those resources.

Our knowledge of the components of human intelligence has come about mostly within the last 25 years. The major sources of this information in this country have been L. L. Thurstone and his associates, the wartime research of psychologists in the United States Air Forces, and more recently the Aptitudes Project[1] at the University of Southern California, now in its tenth year of research on cognitive and thinking abilities. The results from the Aptitudes Project that have gained perhaps the most attention have pertained to creative-thinking abilities. These are mostly novel findings. But to me, the most significant outcome has been the development of a unified theory of human intellect, which organizes the known, unique, or primary intellectual abilities into a single system called the "structure of intellect." It is to this system that I shall devote the major part of my remarks, with very brief mentions of some of the implications for the psychology of thinking and problem solving, for vocational testing, and for education.

The discovery of the components of intelligence has been by means of the experimental application of the method of factor analysis. It is not necessary for you to know anything about the theory or method of factor analysis in order to follow the discussion of the components. I should like to say, however, that factor analysis has no connection with or resemblance to psychoanalysis. A positive statement would be more helpful, so I will say that each intellectual component or factor is a unique ability that is needed to do well in a certain class of tasks or tests. As a general principle we find that certain individuals do well in the tests of a certain class, but they may do poorly in the tests of another class. We conclude that a factor has certain properties from the features that the tests of a class have in common. I shall give you very soon a number of examples of tests, each representing a factor.

---

[1] Under Contract N6onr-23810 with the Office of Naval Research (Personnel and Training Branch).

## The Structure of Intellect

Although each factor is sufficiently distinct to be detected by factor analysis, in very recent years it has become apparent that the factors themselves can be classified because they resemble one another in certain ways. One basis of classification is according to the basic kind of process or operation performed. This kind of classification gives us five major groups of intellectual abilities: factors of cognition, memory, convergent thinking, divergent thinking, and evaluation.

Cognition means discovery or rediscovery or recognition. Memory means retention of what is cognized. Two kinds of productive-thinking operations generate new information from known information and remembered information. In divergent-thinking operations we think in different directions, sometimes searching, sometimes seeking variety. In convergent thinking the information leads to one right answer or to a recognized best or conventional answer. In evaluation we reach decisions as to goodness, correctness, suitability, or adequacy of what we know, what we remember, and what we produce in productive thinking.

A second way of classifying the intellectual factors is according to the kind of material or content involved. The factors known thus far involve three kinds of material or content: the content may be figural, symbolic, or semantic. Figural content is concrete material such as is perceived through the senses. It does not represent anything except itself. Visual material has properties such as size, form, color, location, or texture. Things we hear or feel provide other examples of figural material. Symbolic content is composed of letters, digits, and other conventional signs, usually organized in general systems, such as the alphabet or the number system. Semantic content is in the form of verbal meanings or ideas, for which no examples are necessary.

When a certain operation is applied to a certain kind of content, as many as six general kinds of products may be involved. There is enough evidence available to suggest that, regardless of the combinations of operations and content, the same six kinds of products may be found associated. The six kinds of products are: units, classes, relations, systems, transformations, and implications. So far as we have determined from factor analysis, these are the only fundamental kinds of products that we can know. As such, they may serve as basic classes into which one might fit all kinds of information psychologically.

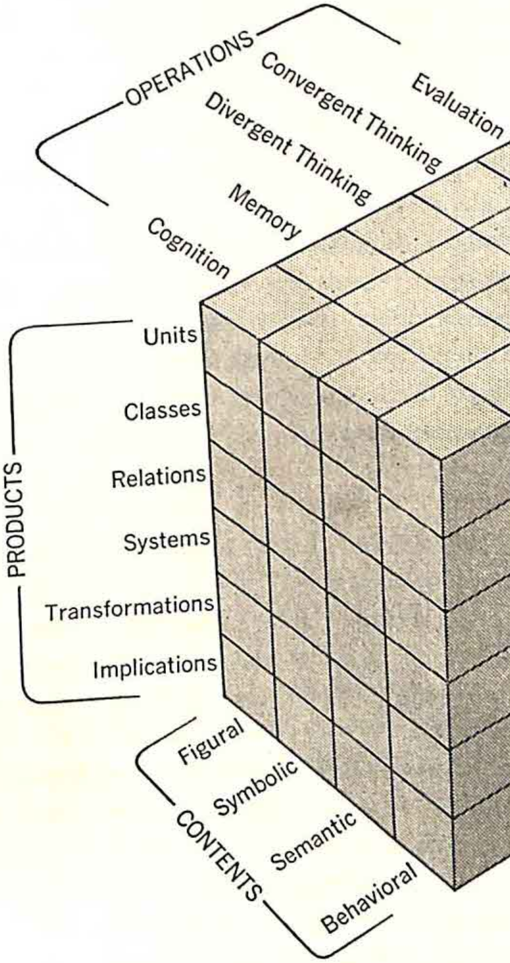


FIG. 1. *A cubical model representing the structure of intellect.*

The three kinds of classifications of the factors of intellect can be represented by means of a single solid model, shown in Figure 1. In this model, which we call the "structure of intellect," each dimension represents one of the modes of variation of the factors.[2] Along one dimension are found the various kinds of operations, along a second one are the various kinds of products, and along the third are various kinds of content. Along the dimension of content a fourth category has been added, its kind of content being designated as "behavioral." This category has been added on a purely theoretical basis to represent the general area sometimes called "social intelligence." More will be said about this section of the model later.

In order to provide a better basis for understanding the model and a better basis for accepting it as a picture of human intellect, I shall do some exploring of it with you systematically, giving some examples of tests. Each cell in the model calls for a certain kind of ability that can be described in terms of operation, content, and product, for each cell is at the intersection of a unique combination of kinds of operation, content, and product. A test for that ability would have the same three properties. In our exploration of the model, we shall take one vertical layer at a time, beginning with the front face. The first layer provides us with a matrix of 18 cells (if we ignore the behaviorial column for which there are as yet no known factors) each of which should contain a cognitive ability.

## The Cognitive Abilities

We know at present the unique abilities that fit logically into 15 of the 18 cells for cognitive abilities. Each row presents a triad of similar abilities, having a single kind of product in common. The factors of the first row are concerned with the knowing of units. A good test of the ability to cognize figural units is the Street Gestalt Completion Test. In this test, the recognition of familiar pictured objects in silhouette form is made difficult for testing purposes by blocking out parts of those objects. There is another factor that is known to involve the perception of auditory figures—in the form of melodies, rhythms, and speech sounds—and still another factor involving kinesthetic forms. The presence of three factors in one cell (they are conceivably distinct abilities, although this has not been tested) suggests that more generally, in the figural column, at least, we should expect to find more than one ability. A fourth dimension pertaining to variations in sense modality may thus apply in connection with figural content. The model could be extended in this manner if the facts call for such an extension.

The ability to cognize symbolic units is measured by tests like the following:

Put vowels in the following blanks to make real words:

P__W__R

M__RV__L

C__RT__N

---

[2] For an earlier presentation of the concept, see Guilford (1956).

Rearrange the letters to make real words:

R A C I H
T V O E S
K L C C O

The first of these two tests is called Disemvoweled Words, and the second Scrambled Words.

The ability to cognize semantic units is the well-known factor of verbal comprehension, which is best measured by means of a vocabulary test, with items such as:

GRAVITY means ________________
CIRCUS means ________________
VIRTUE means ________________

From the comparison of these two factors it is obvious that recognizing familiar words as letter structures and knowing what words mean depend upon quite different abilities.

For testing the abilities to know classes of units, we may present the following kinds of items, one with symbolic content and one with semantic content:

Which letter group does not belong?
XECM PVAA QXIN VTRO

Which object does not belong?
clam tree oven rose

A figural test is constructed in a completely parallel form, presenting in each item four figures, three of which have a property in common and the fourth lacking that property.

The three abilities to see relationships are also readily measured by a common kind of test, differing only in terms of content. The well-known analogies test is applicable, two items in symbolic and semantic form being:

JIRE : KIRE : : FORA : KORE KORA LIRE GORA GIRE
poetry : prose : : dance : music walk sing talk jump

Such tests usually involve more than the ability to cognize relations, but we are not concerned with this problem at this point.

The three factors for cognizing systems do not at present appear in tests so closely resembling one another as in the case of the examples just given. There is nevertheless an underlying common core of logical similarity. Ordinary space tests, such as Thurstone's Flags, Figures, and Cards or Part V (Spatial Orientation) of the Guilford-Zimmerman Aptitude Survey (GZAS), serve in the figural column. The system involved is an order or arrangement of objects in space. A system that uses symbolic elements is illustrated by the Letter Triangle Test, a sample item of which is:

```
          ___
        d     ___
     b     e      ___
  a     c     f      ?
                    ___
```

What letter belongs at the place of the question mark?

The ability to understand a semantic system has been known for some time as the factor called general reasoning. One of its most faithful indicators is a test composed of arithmetic-reasoning items. That the phase of understanding only is important for measuring this ability is shown by the fact that such a test works even if the examinee is not asked to give a complete solution; he need only show that he structures the problem properly. For example, an item from the test Necessary Arithmetical Operations simply asks what operations are needed to solve the problem:

A city lot 48 feet wide and 149 feet deep costs $79,432. What is the cost per square foot?

A. add and multiply
B. multiply and divide
C. subtract and divide
D. add and subtract
E. divide and add

Placing the factor of general reasoning in this cell of the structure of intellect gives us some new conceptions of its nature. It should be a broad ability to grasp all kinds of systems that are conceived in terms of verbal concepts, not restricted to the understanding of problems of an arithmetical type.

Transformations are changes of various kinds, including modifications in arrangement, organization, or meaning. In the figural column for the transformations row, we find the factor known as visualization. Common measuring instruments for this factor are the surface-development tests, and an example of a different kind is Part VI (Spatial Visualization) of the GZAS. A test of the ability to make transformations of meaning, for the factor in the semantic column, is called Similarities. The examinee is asked to state several ways in which two objects, such as an apple and an orange, are alike. Only by shifting the meanings of both is the examinee able to give many responses to such an item.

In the set of abilities having to do with the cognition of implications, we find that the individual goes beyond the information given, but not to the extent of what might be called drawing conclusions. We may say that he extrapolates. From the given information he expects or foresees certain consequences, for example. The two factors found in this row of the cognition matrix were first called "foresight" factors. Foresight in connection with figural material can be tested by means of paper-and-pencil mazes. Foresight in connection with ideas, those pertaining to events, for example, is indicated by a test such as Pertinent Questions:

> In planning to open a new hamburger stand in a certain community, what four questions should be considered in deciding upon its location?

The more questions the examinee asks in response to a list of such problems, the more he evidently foresees contingencies.

## The Memory Abilities

The area of memory abilities has been explored less than some of the other areas of operation, and only seven of the potential cells of the memory matrix have known factors in them. These cells are restricted to three rows: for units, relations, and systems. The first cell in the memory matrix is now occupied by two factors, parallel to two in the corresponding cognition matrix: visual memory and auditory memory. Memory for series of letters or numbers, as in memory-span tests, conforms to the conception of memory for symbolic units. Memory for the ideas in a paragraph conforms to the conception of memory for semantic units.

The formation of associations between units, such as visual forms, syllables, and meaningful words, as in the method of paired associates, would seem to represent three abilities to remember relationships involving three kinds of content. We know of two such abilities, for the symbolic and semantic columns. The memory for known systems is represented by two abilities very recently discovered (Christal, 1958). Remembering the arrangement of objects in space is the nature of an ability in the figural column, and remembering a sequence of events is the nature of a corresponding ability in the semantic column. The differentiation between these two abilities implies that a person may be able to say where he saw an object on a page, but he might not be able to say on which of several pages he saw it after leafing through several pages that included the right one. Considering the blank rows in the memory matrix, we should expect to find abilities also to remember classes, transformations, and implications, as well as units, relations, and systems.

## The Divergent-Thinking Abilities

The unique feature of divergent production is that a *variety* of responses is produced. The product is not completely determined by the given information. This is not to say that divergent thinking does not come into play in the total process of reaching a unique conclusion, for it comes into play wherever there is trial-and-error thinking.

The well-known ability of word fluency is tested by asking the examinee to list words satisfying a specified letter requirement, such as words beginning with the letter "s" or words ending in "-tion." This ability is now regarded as a facility in divergent production of symbolic units. The parallel semantic ability has been known as ideational fluency. A typical test item calls for listing objects that are round and edible. Winston Churchill must have possessed this ability to a high degree. Clement Attlee is reported to have said about him recently that, no matter what problem came up, Churchill always seemed to have about ten ideas. The trouble was, Attlee continued, he did not know which was the good one. The last comment implies some weakness in one or more of the evaluative abilities.

The divergent production of class ideas is believed to be the unique feature of a factor called "spontaneous flexibility." A typical test instructs the examinee to list all the uses he can think of for a common brick, and he is given eight minutes. If his responses are: build a house, build a barn, build a garage, build a school, build a church, build a chimney, build a walk, and build a barbecue, he would earn a fairly high score for ideational fluency but a very low score for spontaneous flexibility, because all these uses fall into the same class. If another person said: make a door stop, make a paper weight, throw it at a dog, make a bookcase, drown a cat, drive a nail, make a red powder, and use for baseball bases, he would also receive a high score for flexibility. He has gone frequently from one class to another.

A current study of unknown but predicted divergent-production abilities includes testing whether there are also figural and symbolic abilities to produce multiple classes. An experimental figural test presents a number of figures that can be classified in groups of three in various ways, each figure being usable in more than one class. An experimental symbolic test presents a few numbers that are also to be classified in multiple ways.

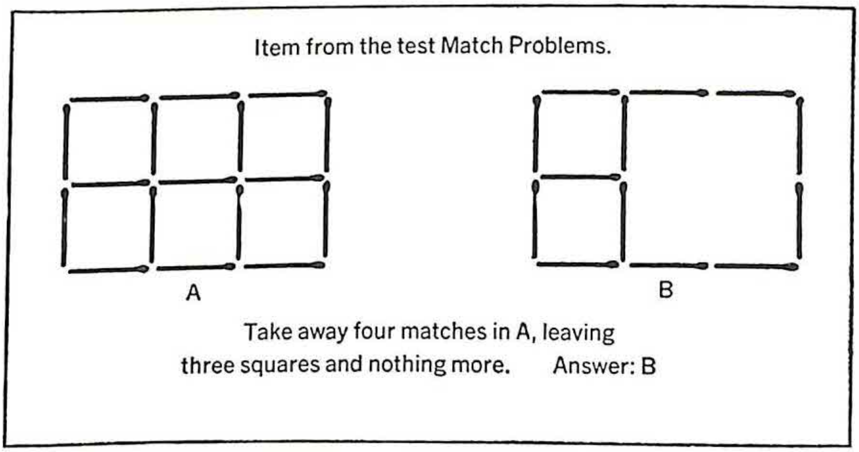
A unique ability involving relations is called "associational fluency." It calls for the production of a variety of things related in a specified way to a given thing. For example, the examinee is asked to list words meaning about the same as "good" or to list words meaning about the opposite of "hard." In these instances the response produced is to complete a relationship, and semantic content is involved. Some of our present experimental tests call for the production of varieties of relations, as such, and involve figural and symbolic content also. For example, given four small digits, in how many ways can they be related in order to produce a sum of eight?

One factor pertaining to the production of systems is known as expressional fluency. The rapid formation of phrases or sentences is the essence of certain tests of this factor. For example, given the initial letters:

W__________ c__________ e__________ n__________

with different sentences to be produced, the examinee might write "We can eat nuts" or "Whence came Eve Newton?" In interpreting the factor, we regard the sentence as a symbolic system. By analogy, a figural system would be some kind of organization of lines and other elements, and a semantic system would be in the form of a verbally stated problem or perhaps something as complex as a theory.

In the row of the divergent-production matrix devoted to transformations, we find some very interesting factors. The one called "adaptive flexibility" is now recognized as belonging in the figural column. A faithful test of it has been Match Problems. This is based upon the common game that uses squares, the sides of which are formed by match sticks. The examinee is told to take away a given number of matches to leave a stated number of squares with nothing left over. Nothing is said about the sizes of the squares to be left. If the examinee imposes upon himself the restriction that the squares that he leaves must be of the same size, he will fail in his attempts to do items like that in Figure 2. Other odd kinds of solutions are introduced in other items, such as overlapping squares and squares within squares, and so on. In another variation of Match Problems the examinee is told to produce two or more solutions for each problem.


Item from the test Match Problems.

Take away four matches in A, leaving three squares and nothing more. Answer: B

FIG. 2. *A sample item from the test Match Problems. The problem in this item is to take away four matches and leave three squares. The solution is given.*

A factor that has been called "originality" is now recognized as adaptive flexibility with semantic material, where there must be a shifting of meanings. The examinee must produce the shifts or changes in meaning and so come up with novel, unusual, clever, or farfetched ideas. The Plot Titles Test presents a short story, the examinee being told to list as many appropriate titles as he can to head the story. One story is about a missionary who has been captured by cannibals in Africa. He is in the pot and about to be boiled when a princess of the tribe obtains a promise for his release if he will become her mate. He refuses and is boiled to death.

In scoring the test, we separate the responses into two categories, clever and nonclever. Examples of nonclever responses are: African Death, Defeat of a Princess, Eaten by Savages, The Princess, The African Missionary, In Darkest Africa, and Boiled by Savages. These titles are appropriate but commonplace. The number of such responses serves as a score for ideational fluency. Examples of clever responses are: Pot's Plott, Potluck Dinner, Stewed Parson, Goil or Boil, A Mate Worse Than Death, He Left a Dish for a Pot, Chaste in Haste, and A Hot Price for Freedom. The number of clever responses given by an examinee is his score for originality, or the divergent production of semantic transformations.

Another test of originality presents a very novel task so that any acceptable response is unusual for the individual. In the Symbol Production Test the examinee is to produce a simple symbol to stand for a noun or a verb in each short sentence, in other words to invent something like pictographic symbols. Still another test of originality asks for writing the "punch lines" for cartoons, a task that almost automatically challenges the examinee to be clever. Thus, quite a variety of tests offer approaches to the measurement of originality, including one or two others that I have not mentioned.

Abilities to produce a variety of implications are assessed by tests calling for elaboration of given information. A figural test of this type provides the examinee with a line or two, to which he is to add other lines to produce an object. The more lines he adds, the greater his score. A semantic test gives the examinee the outlines of a plan to which he is to respond by stating all the details he can think of to make the plan work. A new test we are trying out in the symbolic area presents two simple equations such as $B - C = D$ and $z = A + D$. The examinee is to make as many other equations as he can from this information.

## The Convergent-Production Abilities

Of the 18 convergent-production abilities expected in the three content columns, 12 are now recognized. In the first row, pertaining to units, we have an ability to name figural properties (forms or colors) and an ability to name abstractions (classes, relations, and so on). It may be that the ability in common to the speed of naming forms and the speed of naming colors is not appropriately placed in the convergent-thinking matrix. One might expect that the thing to be produced in a test of the convergent production of figural units would be in the form of figures rather than words. A better test of such an ability might somehow specify the need for one particular object, the examinee to furnish the object.

A test for the convergent production of classes (Word Grouping) presents a list of 12 words that are to be classified in four, and only four, meaningful groups, no word to appear in more than one group. A parallel test (Figure Concepts Test) presents 20 pictured real objects that are to be grouped in meaningful classes of two or more each.

Convergent production having to do with relationships is represented by three known factors, all involving the "education of correlates," as Spearman called it. The given information includes one unit and a stated relation, the examinee to supply the other unit. Analogies tests that call for completion rather than a choice between alternative answers emphasize this kind of ability. With symbolic content such an item might read:

pots stop bard drab rats ?

A semantic item that measures education of correlates is:

The absence of sound is ______________.

Incidentally, the latter item is from a vocabulary-completion test, and its relation to the factor of ability to produce correlates indicates how, by change of form, a vocabulary test may indicate an ability other than that for which vocabulary tests are usually intended, namely, the factor of verbal comprehension.

Only one factor for convergent production of systems is known, and it is in the semantic column. It is measured by a class of tests that may be called ordering tests. The examinee may be presented with a number of events that ordinarily have a best or most logical order, the events being presented in scrambled order. The presentation may be pictorial, as in the Picture Arrangement Test, or verbal. The pictures may be taken from a cartoon strip. The verbally presented events may be in the form of the various steps needed to plant a new lawn. There are undoubtedly other kinds of systems than temporal order that could be utilized for testing abilities in this row of the convergent-production matrix.

In the way of producing transformations of a unique variety, we have three recognized factors, known as redefinition abilities. In each case, redefinition involves the changing of functions or uses of parts of one unit and giving them new functions or uses in some new unit. For testing the ability of figural redefinition, a task based upon the Gottschaldt figures is suitable. Figure 3 shows the kind of item for such a test. In recognizing the simpler figure within the structure of a more complex figure, certain lines must take on new roles.

In terms of symbolic material, the following sample items will illustrate how groups of letters in given words must be readapted to use in other words. In the test Camouflaged Words, each sentence contains the name of a sport or game:

I did not know that he was ailing.
To beat the Hun, tin goes a long way.

For the factor of semantic redefinition, the Gestalt Transformation Test may be used. A sample item reads:

From which object could you most likely make a needle?
- A. a cabbage
- B. a splice
- C. a steak
- D. a paper box
- E. a fish

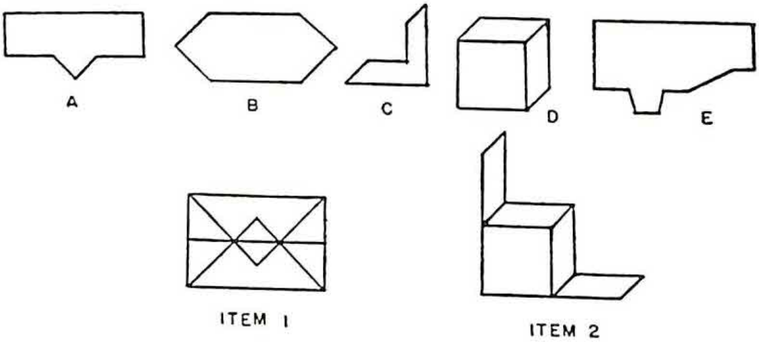


FIG. 3. *Sample items from a test Hidden Figures, based upon the Gottschaldt figures. Which of the simpler figures is concealed within each of the two more complex figures?*

The convergent production of implications means the drawing of fully determined conclusions from given information. The well-known factor of numerical facility belongs in the symbolic column. For the parallel ability in the figural column, we have a test known as Form Reasoning, in which rigorously defined operations with figures are used. For the parallel ability in the semantic column, the factor sometimes called "deduction" probably qualifies. Items of the following type are sometimes used.

Charles is younger than Robert
Charles is older than Frank
Who is older: Robert or Frank?

### Evaluative Abilities

The evaluative area has had the least investigation of all the operational categories. In fact, only one systematic analytical study has been devoted to this area. Only eight evaluative abilities are recognized as fitting into the evaluation matrix. But at least five rows have one or more factors each, and also three of the usual columns or content categories. In each case, evaluation involves reaching decisions as to the accuracy, goodness, suitability, or workability of information. In each row, for the particular kind of product of that row, some kind of criterion or standard of judgment is involved.

In the first row, for the evaluation of units, the important decision to be made pertains to the identity of a unit. Is this unit identical with that one? In the figural column we find the factor long known as "perceptual speed." Tests of this factor invariably call for decisions of identity, for example, Part IV (Perceptual Speed) of the GZAS or Thurstone's Identical Forms. I think it has been generally wrongly thought that the ability involved is that of cognition of visual forms. But we have seen that another factor is a more suitable candidate for this definition and for being in the very first cell of the cognitive matrix. It is parallel to this evaluative ability but does not require the judgment of identity as one of its properties.

In the symbolic column is an ability to judge identity of symbolic units, in the form of series of letters or numbers or of names of individuals.

Are members of the following pairs identical or not:

825170493_______825176493
dkeltvmpa_______dkeltvmpa
C. S. Meyerson_______C. E. Meyerson

Such items are common in tests of clerical aptitude.

There should be a parallel ability to decide whether two ideas are identical or different. Is the idea expressed in this sentence the same as the idea expressed in that one? Do these two proverbs express essentially the same idea? Such tests exist and will be used to test the hypothesis that such an ability can be demonstrated.

No evaluative abilities pertaining to classes have as yet been recognized. The abilities having to do with evaluation where relations are concerned must meet the criterion of logical consistency. Syllogistic-type tests involving letter symbols indicate a different ability than the same type of test involving verbal statements. In the figural column we might expect that tests incorporating geometric reasoning or proof would indicate a parallel ability to sense the soundness of conclusions regarding figural relationships.

The evaluation of systems seems to be concerned with the internal consistency of those systems, so far as we can tell from the knowledge of one such factor. The factor has been called "experiential evaluation," and its representative test presents items like that in Figure 4 asking "What is wrong with this picture?" The things wrong are often internal inconsistencies.

A semantic ability for evaluating transformations is thought to be that known for some time as "judgment." In typical judgment tests, the examinee is asked to tell which of five solutions to a practical problem is most adequate or wise. The solutions frequently involve improvisations, in other words, adaptations of familiar objects to unusual uses. In this way the items present redefinitions to be evaluated.

A factor known first as "sensitivity to problems" has become recognized as an evaluative ability having to do with implications. One test of the factor, the Apparatus Test, asks for two needed improvements with respect to each of several common devices, such as the telephone or the toaster. The Social Institutions Test, a measure of the same factor, asks what things are wrong with each of several institutions, such as tipping or national elections. We may say that defects or deficiencies are implications of an evaluative kind. Another interpretation would be that seeing defects and deficiencies are evaluations of implications to the effect that the various aspects of something are all right.[3]

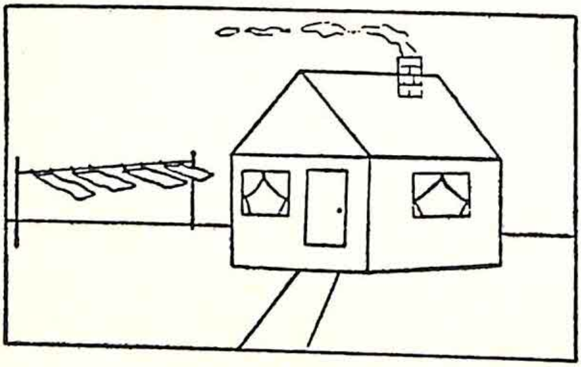


FIG. 4. *A sample item from the test Unusual Details. What two things are wrong with this picture?*

## Some Implications of the Structure of Intellect

### For Psychological Theory

Although factor analysis as generally employed is best designed to investigate ways in which individuals differ from one another, in other words, to discover traits, the results also tell us much about how individuals are alike. Consequently, information regarding the factors and their interrelationships gives us understanding of functioning individuals. The five kinds of intellectual abilities in terms of operations may be said to represent five ways of functioning. The kinds of intellectual abilities distinguished according to varieties of test content and the kinds of abilities distinguished according to varieties of products suggest a classification of basic forms of information or knowledge. The kind of organism suggested by this way of looking at intellect is that of an agency for dealing with information of various kinds in various ways. The concepts provided by the distinctions among the intellectual abilities and by their classifications may be very useful in our future investigations of learning, memory, problem solving, invention, and decision making, by whatever method we choose to approach those problems.

---

[3] For further details concerning the intellectual factors, illustrative tests, and the place of the factors in the structure of intellect; see Guilford (1959).

### For Vocational Testing

With about 50 intellectual factors already known, we may say that there are at least 50 ways of being intelligent. It has been facetiously suggested that there seem to be a great many more ways of being stupid, unfortunately. The structure of intellect is a theoretical model that predicts as many as 120 distinct abilities, if every cell of the model contains a factor. Already we know that two cells contain two or more factors each, and there probably are actually other cells of this type. Since the model was first conceived, 12 factors predicted by it have found places in it. There is consequently hope of filling many of the other vacancies, and we may eventually end up with more than 120 abilities.

The major implication for the assessment of intelligence is that to know an individual's intellectual resources thoroughly we shall need a surprisingly large number of scores. It is expected that many of the factors are intercorrelated, so there is some possibility that by appropriate sampling we shall be able to cover the important abilities with a more limited number of tests. At any rate, a multiple-score approach to the assessment of intelligence is definitely indicated in connection with future vocational operations.

Considering the kinds of abilities classified as to content, we may speak roughly of four kinds of intelligence. The abilities involving the use of figural information may be regarded as "concrete" intelligence. The people who depend most upon these abilities deal with concrete things and their properties. Among these people are mechanics, operators of machines, engineers (in some aspects of their work), artists, and musicians.

In the abilities pertaining to symbolic and semantic content, we have two kinds of "abstract" intelligence. Symbolic abilities should be important in learning to recognize words, to spell, and to operate with numbers. Language and mathematics should depend very much upon them, except that in mathematics some aspects, such as geometry, have strong figural involvement. Semantic intelligence is important for understanding things in terms of verbal concepts and hence is important in all courses where the learning of facts and ideas is essential.

In the hypothesized behavioral column of the structure of intellect, which may be roughly described as "social" intelligence, we have some of the most interesting possibilities. Understanding the behavior of others and of ourselves is largely nonverbal in character. The theory suggests as many as 30 abilities in this area, some having to do with understanding, some with productive thinking about behavior, and some with the evaluation of behavior. The theory also suggests that information regarding behavior is also in the form of the six kinds of products that apply elsewhere in the structure of intellect, including units, relations, systems, and so on. The abilities in the area of social intelligence, whatever they prove to be, will possess considerable importance in connection with all those individuals who deal most with other people: teachers, law officials, social workers, therapists, politicians, statesmen, and leaders of other kinds.

### For Education

The implications for education are numerous, and I have time just to mention a very few. The most fundamental implication is that we might well undergo transformations with respect to our conception of the learner and of the process of learning. Under the prevailing conception, the learner is a kind of stimulus-response device, much on the order of a vending machine. You put in a coin, and something comes out. The machine learns what reaction to put out when a certain coin is put in. If, instead, we think of the learner as an agent for dealing with information, where information is defined very broadly, we have something more analogous to an electronic computor. We feed a computor information; it stores that information; it uses that information for generating new information, either by way of divergent or convergent thinking; and it evaluates its own results. Advantages that a human learner has over a computor include the step of seeking and discovering new information from sources outside itself and the step of programing itself. Perhaps even these steps will be added to computors, if this has not already been done in some cases.

At any rate, this conception of the learner leads us to the idea that learning is discovery of information, not merely the formation of associations, particularly associations in the form of stimulus-response connections. I am aware of the fact that my proposal is rank heresy. But if we are to make significant progress in our understanding of human learning and particularly our understanding of the so-called higher mental processes of thinking, problem solving, and creative thinking, some drastic modifications are due in our theory.

The idea that education is a matter of training the mind or of training the intellect has been rather unpopular, wherever the prevailing psychological doctrines have been followed. In theory, at least, the emphasis has been upon the learning of rather specific habits or skills. If we take our cue from factor theory, however, we recognize that most learning probably has both specific and general aspects or components. The general aspects may be along the lines of the factors of intellect. This is not to say that the individual's status in each factor is entirely determined by learning. We do not know to what extent each factor is determined by heredity and to what extent by learning. The best position for educators to take is that possibly every intellectual factor can be developed in individuals at least to some extent by learning.

If education has the general objective of developing the intellects of students, it can be suggested that each intellectual factor provides a particular goal at which to aim. Defined by a certain combination of content, operation, and product, each goal ability then calls for certain kinds of practice in order to achieve improvement in it. This implies choice of curriculum and the choice or invention of teaching methods that will most likely accomplish the desired results.

Considering the very great variety of abilities revealed by the factorial exploration of intellect, we are in a better position to ask whether any general intellectual skills are now being neglected in education and whether appropriate balances are being observed. It is often observed these days that we have fallen down in the way of producing resourceful, creative graduates. How true this is, in comparison with other times, I do not know. Perhaps the deficit is noticed because the demands for inventiveness are so much greater at this time. At any rate, realization that the more conspicuously creative abilities appear to be concentrated in the divergent-thinking category, and also to some extent in the transformation category, we now ask whether we have been giving these skills appropriate exercise. It is probable that we need a better balance of training in the divergent-thinking area as compared with training in convergent thinking and in critical thinking or evaluation.

The structure of intellect as I have presented it to you may or may not stand the test of time. Even if the general form persists, there are likely to be some modifications. Possibly some different kind of model will be invented. Be that as it may, the fact of a multiplicity of intellectual abilities seems well established.

There are many individuals who long for the good old days of simplicity, when we got along with one unanalyzed intelligence. Simplicity certainly has its appeal. But human nature is exceedingly complex, and we may as well face that fact. The rapidly moving events of the world in which we live have forced upon us the need for knowing human intelligence thoroughly. Humanity's peaceful pursuit of happiness depends upon our control of nature and of our own behavior; and this, in turn, depends upon understanding ourselves, including our intellectual resources.

## REFERENCES

Christal, R. E. Factor analytic study of visual memory. *Psychol. Monogr.*, 1958, 72, No. 13 (Whole No. 466).

Guilford, J. P. The structure of intellect. *Psychol. Bull.*, 1956, 53, 267–293.

Guilford, J. P. *Personality*. New York: McGraw-Hill, 1959.

# 5

# Intelligence at Adulthood: A Twenty-Five Year Follow-Up — *Katherine P. Bradway and Clare W. Thompson*

*This article reports a longitudinal study involving 111 subjects who had been tested first in preschool and adolescent years with the Stanford-Binet and who later, in adulthood, were administered the Stanford-Binet and the Wechsler Adult Intelligence Scale. Some interesting findings related to sex differences and to the level of educational attainment are noted. Although the authors' analysis of results indicates a certain degree of stability over a 25-year period, the pattern of individual differences, measured in relative amounts for different intellectual abilities, indicates that mental growth is not the same in all areas. This study reinforces the notion, presented in the previous article, that human intelligence is a highly complex concept; it cannot be meaningfully expressed by a single index.*

ATTEMPTS TO understand human intelligence were based initially on studies of children. In contrast, this paper deals with aspects of intelligence which cannot be studied adequately from data limited to childhood intelligence. Specifically, our approach has been to study as young adults a group of individuals previously tested during the preschool years and during

Reprinted from *Journal of Educational Psychology*, 1962, **53**, 1–14, by permission of the senior author and the American Psychological Association.

adolescence (Bradway 1944, 1945a, 1945b; Bradway & Robinson 1961; Bradway, Thompson, & Cravens 1958).

Binet and Simon developed the original intelligence test to objectify the criteria which educators had been using for many years in assessing the capacities of children for purposes of school placement. The several early American adaptations of this test culminated in the 1916 Stanford-Binet, which provided indices of adult intelligence. The standardization group, however, consisted largely of children; adult indices were based on extrapolation.

The urgency of the needs for screening and assignment created by World War I led to the creation of truly adult intelligence tests but, consistent with these needs, these tests were for group rather than individual administration. Later revisions of individual intelligence tests (e.g., the 1937 and 1960 Stanford and the 1939 Kuhlmann) extended the upper limit by providing more difficult test items. Moreover, somewhat older subjects were included in the standardization groups, although the upper age levels of these groups did not go above the adolescent years into adulthood.

No adequate instrument existed by which the clinician could measure the intelligence of the individual adult until the publication of the Wechsler-Bellevue in 1939. Wechsler's large standardization sample started with children of 12 years and included older age groups up to 65 years. His emphasis on adult intelligence resulted in departure from the use of mental age, a concept which had contributed immeasurably to the interpretation of childhood intelligence, but which introduced confusion in the assessment of adult intelligence. His term "IQ" is not a continuation of Stern's quotient of mental age divided by chronological age; it is more comparable to a standard score. Unlike the revisions of the Binet, all of which had been based on the concept that intellectual abilities reach their peak at from 14 to 16 years of age, indices of intelligence for the first time took account of the continuation of intellectual growth into the adult years. This view of intellectual growth has been increasingly confirmed by more recent studies (Bayley, 1955; Bayley & Oden, 1955; Miner, 1957; Owens, 1953) including the present one. Although Wechsler calls his concept of intelligence global, the structure of the Wechsler-Bellevue (and its later revision, the WAIS) is such that several kinds of tasks are separately measured and separate indices of verbal and nonverbal abilities are provided.

The clinician trying to answer questions about the intelligence of an adult needs information about the unique growth rates of different kinds of abilities. No cross-sectional sampling can provide such information accurately. On the other hand, longitudinal studies pose problems of practice effects and of the attrition which inevitably occurs over a period of time. The clinician is also confronted with the problem of how most meaningfully to express the measured intelligence of any given adult. Should it be an index which compares him with his own age group? One which locates him on his own course of development and decline? Or several which represent an assessment of his various kinds of mental processes? Moreover, each of these requires that procedures for equating means and variability be worked out. Answers to many of these questions are necessary before valid predictions about adult intelligence can be made.

## Procedure and Subjects

Information for some of these answers is provided by the longitudinal study we have recently completed (Bradway & Robinson, 1961; Bradway et al., 1958). A group of 111 adults aged 27 to 32 who had been tested 25 years earlier as part of the 1937 Stanford-Binet standardization sample of preschool children and again during their adolescent years (Bradway 1944, 1945a, 1945b) was administered the Stanford-Binet and WAIS in 1956.[1] The repetition of the Stanford-Binet permits analysis of improvement in the same tasks from adolescence to adulthood; the WAIS data permit comparison of child and adult measures of intelligence of the same individuals. Treatment of the data has included correlations between total scores, mental growth curves, an analysis of items for increase in percent passing from adolescence to adulthood, and an analysis of the predictive value of preschool items. This latter item analysis was based on: correlations between Stanford-Binet (Forms L and M) preschool items and both adolescent Stanford-Binet IQ and adult WAIS verbal and performance quotients, and comparisons between scores on special scales constructed from the several kinds of preschool items and later scores obtained on comparable adult WAIS scales. Because for one subject the WAIS was incomplete and educational attainment was not ascertained, and for another the 1931 Stanford-Binet IQ was invalid, the number of total subjects on which any one analysis is based varies from 109 to 111 depending upon which variables are involved in the analysis. No attempt was made in the two follow-ups to continue the procedure used in the initial examination of testing children within one month of a birthday; a larger deviation in the age distribution was a consequence.

[1] The authors are happy to express their gratitude to Maud A. Merrill and the late Lewis M. Terman for generously having made their standardization data available for follow-up; to Quinn McNemar, consultant and advisor throughout the study; and to Nancy M. Robinson who participated actively in the later stages of the study.

## Results and Discussion

### Constancy of the IQ

A finding of primary interest is that an individual's measured intelligence in relation to others of his same age is demonstrated to be more stable than previous studies suggest. These results are presented in Table 1. It will be noted that the correlation between preschool (1931) intelligence, and adult (1956) intelligence as measured by the full scale WAIS is .64. The measure of preschool intelligence was a combination of the two forms of the Stanford-Binet which became the 1937 Revision, and the measure of adult intelligence was a combination of the verbal and performance scales of the WAIS; thus the indices of intelligence at both ages were based on the most adequate measurements available when the data were collected. The reliability of each measurement was sufficiently high so that correcting for unreliability increases the correlation only slightly: from .64 to .69. The degree of correspondence over a 25-year period is not significantly different from the correlation of .65 (uncorrected) found between preschool and adolescent Stanford-Binet IQs after an interval of 10 years.

TABLE 1. Correlations between Test Scores at the Several Administrations

| Test | 1941 S-B (L) | 1956 S-B (L) | 1956 WAIS Full | 1956 WAIS Verbal | 1956 WAIS Performance |
|---|---|---|---|---|---|
| 1931 S-B (L and M) | .65 | .59 | .64 | .60 | .54 |
| 1941 S-B (L) | | .85 | .80 | .81 | .51 |
| 1956 S-B (L) | | | .83 | .89 | .46 |
| 1956 WAIS Full | | | | .87 | .84 |
| 1956 WAIS Verbal | | | | | .59 |

Also of note from Table 1 is the correlation of .80 between adolescent (1941) Stanford-Binet and adult (1956) WAIS. Although this is somewhat lower than the correlation of .85 between the adolescent and adult administrations of the Stanford-Binet, neither correlation is significantly different from the correlation of .83 between Stanford-Binet and WAIS administered to our subjects on the same day when they were adults. Apparently whatever the Stanford-Binet measures at preschool and at adolescence is still being measured by the WAIS at adulthood.

A comparison of the columns of Table 1 shows a close similarity, both in absolute amount and in trend, between the 1956 verbal WAIS and the 1956 full scale WAIS when each is correlated with the Stanford-Binet administered at the three age levels. The performance WAIS correlations, on the other hand, are both lower in value and opposite in trend. The reason for this reversal in trend needs further study to determine whether the results reflect an increase in specialization of intelligence with age or are due to an increase in verbal content of the Stanford-Binet at the higher levels.

### Growth and Decline of Intellectual Abilities

Our data contribute to the accumulating evidence that intellectual potential does not cease to increase at 16 years in contradiction to the conclusions of Terman and others. Table 2 summarizes our data for comparison of obtained IQs. In view of the conformance between the 1931 and 1941 Stanford-Binet IQs and between the standard deviations at the three ages, the fact that the 1956 Stanford-Binet IQ exceeds the 1941 Stanford-Binet IQ by a statistically significant amount is interpreted as a consequence of intellectual growth in these subjects beyond 16 years. The increase of 11 points in IQ corresponds to a gain in mental age score of 20 months beyond expectancy. The finding of a continuation of mental growth into adulthood is consistent with the findings of Bayley (1955, 1957) and others (Bayley & Oden, 1955; Miner, 1957; Owens, 1953). The implications of the difference between present and previously held conclusions regarding mental growth into adulthood deserve consideration. Bayley (1955, p. 817) suggests alternative explanations including the new generations' having more opportunity to develop their intellectual capacities than did their parents. Inhelder and Piaget (1958, p. 337) have pointed out the probability that Greek children were behind our own in growth of logical thinking and suggest that the age at

TABLE 2. Means and Standard Deviations of the Two Tests at the Three Administrations

| Test | 1931 *M* | 1931 *SD* | 1941 *M* | 1941 *SD* | 1956 *M* | 1956 *SD* |
|---|---|---|---|---|---|---|
| Chronological age | 4.0 | 1.1 | 13.6 | 1.4 | 29.5 | 1.5 |
| S-B IQ | 112.8[a] | 15.9[a] | 112.3[b] | 16.4[b] | 123.6[b] | 15.0[b] |
| WAIS IQ | | | | | | |
| Full | | | | | 108.9 | 11.0 |
| Verbal | | | | | 110.4 | 13.3 |
| Performance | | | | | 105.0 | 11.5 |

[a] Composite Form L and Form M.
[b] Form L.

which formal thinking begins "may be a product of progressive acceleration of individual development under the influence of education."

The method of computing WAIS IQs assumes continuation of intellectual growth up to 34 years. The mean WAIS IQ of our group is lower in numerical value than expectancy based on the mean 1941 Stanford-Binet IQ. It will be noted, however, that the WAIS mean IQ is approximately one *SD* above 100, whereas the 1941 Stanford-Binet mean IQ is less than one *SD* above 100; so that in terms of relative deviations, the WAIS mean IQ is higher than the Stanford-Binet IQ, although not to a statistically significant degree. The verbal half of the WAIS approaches the 1941 Stanford-Binet results more closely than does the full scale, and the performance half less so.

The addition of the present data to those previously collected for this group of subjects gives us a third segment of the mental growth curve. Thus we can plot mental age against chronological age for the years 2 through 5, the years 11 through 15, and now the years 27 through 31. Two subjects were younger and one was older than these years and are omitted from this section. Table 3 summarizes the current data for the Stanford-Binet.

TABLE 3. Stanford-Binet Scores for Adults by Year of Age

| 1956 Chronological Age | *N* | S-B Mean 1956 IQ | S-B Mean 1956 MA |
|---|---|---|---|
| 27.0–27.9 | 16 | 121.3 | 18.2 |
| 28.0–28.9 | 25 | 125.5 | 18.9 |
| 29.0–29.9 | 23 | 126.8 | 19.0 |
| 30.0–30.9 | 20 | 123.5 | 18.5 |
| 31.0–31.9 | 24 | 121.0 | 18.1 |

Analysis of variance was applied to these data. The resulting *F* ratio was .594 for 4 and 103 degrees of freedom, whereas 2.45 would be required for the .05 level of significance. Of course, should the drop from age 29 to age 31 continue in direction and amount, the critical point of this mental growth curve would have been established, but there is no way to predict this from the present data.

Before we interpret what this lack of change during this 5-year period may mean, we must consider the possible ceiling effect. The 1937 Stanford-Binet has insufficient top for adults and therefore this flattening of the curve might be an artifact. To examine this possibility, a similar analysis was made of the results on the WAIS for these subjects, inasmuch as a low ceiling is not a limitation of that test. One of the subjects had a partial WAIS only and is omitted for this reason, reducing the degrees of freedom to 4 and 102, but not changing the value of 2.45 for the .05 level of significance against which

TABLE 4. WAIS Scores for Adults by Year of Age

| 1956 Chronological Age | *N* | Mean 1956 WAIS Standard Scores<br>Full | Verbal | Performance |
|---|---|---|---|---|
| 27.0–27.9 | 16 | 119 | 67 | 51 |
| 28.0–28.9 | 24 | 127 | 72 | 56 |
| 29.0–29.9 | 23 | 129 | 73 | 55 |
| 30.0–30.9 | 20 | 126 | 72 | 54 |
| 31.0–31.9 | 24 | 119 | 70 | 50 |

the results are compared. Table 4 summarizes the data for the WAIS. Analysis of variance for these three scores again shows the variation with chronological age not to approach significance. The *F* ratios are, respectively, 1.40, 1.01, and 1.62.

Further study to determine the role of ceiling effect was made by analyzing only those cases showing no or few passes at the Superior Adult III level of the Stanford-Binet. Also, because our group of subjects is somewhat above average, a separate study was made of those cases clustering around IQ 100. These additional studies, however, served only to substantiate the results just reported. The point at which mental growth as measured by these instruments terminates and at which the curve assumes a downward direction is not established by these data.

### Sex Differences

Tests such as the Stanford-Binet are intentionally constructed to eliminate or to minimize sex differences. Consequently most investigators, including the present ones in earlier studies, have given little attention to this variable. It is in this area, however, that some of the most cogent findings of the present study emerge. As shown in Table 5, on all three administrations of the Stanford-Binet, female subjects achieved a higher mean score than did males. On the first two administrations, this difference was significant beyond the .01 level. On the 1956 testing, it did not reach significance. The males increased their scores from 1941 to 1956 to a greater extent than did the women. For the males, this increase was 13.7 and for the females it was 9.6; the difference of 4.1 is significant beyond the .01 level. Both females and males show a negative correlation between the 1941 Stanford-Binet IQ and the gain in IQ from 1941 to 1956, but the one for females is larger. For men this correlation is −.23 and for women −.67; the difference between these correlations is significant beyond the .01 level. Dividing the male and female groups according to whether they stopped their education at high school graduation or went beyond does not change the findings. The obtained correlations for males are −.28 and −.19 (not significantly different); and

TABLE 5. Sex Differences on the Stanford-Binet

| Test | Male | Female | Difference |
|---|---|---|---|
| 1931 L-M | | | |
| *N* | 51 | 59 | |
| *M* | 108.5 | 115.8 | −7.3 |
| *SD* | 14.4 | 13.3 | |
| *CR*° | | | 2.79 |
| 1941 L | | | |
| *N* | 52 | 59 | |
| *M* | 107.9 | 116.1 | −8.2 |
| *SD* | 19.0 | 13.7 | |
| *CR*° | | | 2.58 |
| 1956 L | | | |
| *N* | 52 | 59 | |
| *M* | 121.6 | 125.7 | −4.1 |
| *SD* | 19.6 | 10.6 | |
| *CR*° | | | 1.36 |

° A value of 2.57 is significant beyond the .01 level and a value of 1.96 beyond the .05 level.

for females −.70 and −.65 (not significantly different). The decrease in standard deviation for females from 1941 to 1956 as shown in Table 5 reflects that these findings are due to the tendency of women with IQs which are high at adolescence not to increase their scores by adulthood as much as does the rest of the group.

A study of the individuals scoring high (from 130 to 150 IQ) on the 1956 Stanford-Binet yields further information. Females in this group had always scored high relative to the group: mean IQ of 125.9 in 1931, 127.8 in 1941, and 137.5 in 1956. For the males, however, becoming a member of this group represented a gain from an earlier lower point with most of the gain achieved by adolescence: mean IQ of 115.9 in 1931, 126.2 in 1941, and 140.7 in 1956. (It must be remembered that these 1956 Stanford-Binet IQs are, on the average, 11 points high, since they do not include a correction for the fact that there was a mean IQ increase of 11 points for the total sample.) Slightly but not significantly more women than men in this subgroup stopped education with high school graduation. Of those who started college, not a single man failed to finish, whereas one-third of the women dropped out before finishing. This leaves us with the familiar chicken-egg dilemma. Did the women drop out of college because their relative brightness as compared with men decreased—their promise not being realized in the same way as was that of the men? Or did the earlier termination of their academic education make the intelligence tests administered to them as adults relatively more difficult for them?

An analysis of the relation between IQ at adolescence and continuation of education for the total group was made. There is no significant difference for either sex between mean 1941 IQ of those who stopped their education with graduation from high school and the mean 1941 IQ of those who attended college but did not graduate (males: 100.7 and 102.5; females: 113.8 and 113.0). The mean 1941 IQ, however, of both the 17 males and the 14 females who graduated from college is significantly higher (124.6 and 125.0, respectively) than for those who did not graduate from college. Gain in IQ, on the other hand, is not consistently related to educational attainment. Mean IQ gains for males who did not go beyond high school graduation, who entered college but did not graduate, and who graduated from college were 14.7, 11.7, and 11.8, respectively. Corresponding mean gains for females were 8.2, 12.3, and 9.3, respectively. The differences between the mean gains for either sex are not large enough to be statistically significant with the small number of subjects involved (males: 22, 13, and 17 for successive educational attainments; females: 22, 22, and 14).

Because other investigators of adult intelligence have not reported sex differences of the nature found here, we cannot look beyond our own data for the answers to some of the questions which are posed. This seems to open up an area worthy of further study designed explicitly to explore this problem.

## Increases in Abilities from Adolescence to Adulthood

With the establishment of the fact that there is increase in intelligence scores between adolescence and adulthood, one becomes interested in identifying the kinds of abilities which increase most. Within the range of abilities covered by the adult levels of the Stanford-Binet, it was possible to shed light on this question by comparing the proportion of our group passing each item as adolescents and as adults. The percentages of the present sample passing each item were calculated for the 1941 and 1956 results. Items occurring below a subject's basal age were assumed to have been passed; those occurring above a ceiling were assumed to have been failed. Since nearly all subjects passed in 1956 all those items through Level 14, it was possible to make meaningful comparisons for the four adult levels only. Table 6 lists each of the 26 items at these levels, arranged according to the magnitude of the increase in percent passing over the 15-year period. The final column in the table classifies items by "type," following a system suggested by Davis (1941), into rote memory (M), reasoning (R), and vocabulary (V). The reasoning subscale has been further divided by Robinson[2] into those items which appear to involve mental manipulation of practical, concrete concepts (RP) versus those dealing with more abstract concepts

[2] Nancy M. Robinson made several contributions to the treatment of data for this section of the study.

TABLE 6. Items of Stanford-Binet (L) Adult Levels by Magnitude of Change in Percent Passing from Adolescence to Adulthood

| Rank | Location | Item | Percent Passing[a] 1941 | 1956 | Increase | Chi square[b] | Type[c] |
|---|---|---|---|---|---|---|---|
| 1 | AA-5 | Proverbs I | 22.5 | 81.1 | 58.6 | 76.26 | RA |
| 2 | SAI-1 | Vocabulary | 35.1 | 92.8 | 57.7 | 80.07 | V |
| 3 | SAII-1 | Vocabulary | 13.5 | 67.6 | 54.1 | 67.28 | V |
| 4 | SAI-6 | Essential Similarities | 31.5 | 82.9 | 51.4 | 52.42 | RA |
| 5 | SAI-5 | Sentence Building | 30.6 | 77.5 | 46.9 | 49.04 | V |
| 6.5 | AA-6 | Ingenuity | 30.6 | 73.9 | 43.3 | 38.01 | RP |
| 6.5 | SAIII-1 | Vocabulary | 1.8 | 45.0 | 43.2 | 52.84 | V |
| 8 | AA-3 | Difference between Abstract Words | 49.5 | 91.9 | 42.4 | 48.05 | V |
| 9 | SAIII-3 | Opposite Analogies | 8.1 | 48.7 | 40.6 | 44.89 | V |
| 10.5 | AA-2 | Codes | 37.8 | 77.5 | 39.7 | 35.73 | RP |
| 10.5 | SAII-5 | Reconciliation of Opposites | 29.7 | 69.4 | 39.7 | 34.89 | RA |
| 12 | AA-1 | Vocabulary | 59.4 | 98.2 | 38.8 | 49.41 | V |
| 13 | SAII-4 | Proverbs II | 6.3 | 44.1 | 37.8 | 42.81 | RA |
| 14 | AA-7 | Memory for Sentences V | 36.9 | 72.1 | 35.2 | 27.63 | M |
| 15 | SAII-2 | Finding Reasons II | 14.4 | 47.8 | 33.4 | 28.68 | RA |
| 16 | AA-8 | Reconciliation of Opposites | 54.9 | 86.5 | 31.6 | 26.64 | RA |
| 17 | AA-4 | Arithmetic Reasoning | 41.4 | 72.1 | 30.7 | 21.22 | RP |
| 18 | SAIII-2 | Orientation: Directions II | 14.4 | 41.4 | 27.0 | 20.15 | RP |
| 19 | SAI-3 | Minkus Completion | 24.3 | 50.5 | 26.2 | 16.18 | V |
| 20 | SAI-4 | Repeating 6 Digits Reversed | 29.7 | 54.1 | 24.4 | 13.45 | M |
| 21 | SAIII-5 | Reasoning | 3.6 | 24.3 | 20.7 | 19.82 | RP |
| 22 | SAII-3 | Repeating 8 Digits | 17.1 | 35.1 | 18.0 | 9.34 | M |
| 23 | SAIII-6 | Repeating 9 Digits | 3.6 | 16.2 | 12.6 | 9.89 | M |
| 24 | SAIII-4 | Paper Cutting II | 11.7 | 22.5 | 10.8 | 3.57 | RP |
| 25 | SAI-2 | Enclosed Box Problem | 50.4 | 55.9 | 5.5 | 0.62 | RP |
| 26 | SAII-6 | Repeating Thought of Passage | 24.3 | 27.9 | 3.6 | 0.09 | M |

[a] $N = 111$.

[b] A chi square value of 10.83 is significant beyond the .001 level, 6.64 beyond the .01 level, and 3.84 at the .05 level. Hence all but the last three items of this table show a significant increase in percent passing from adolescence to adulthood.

[c] These refer to the classification of Davis and Robinson described previously.

(RA). One might wonder whether those items having the lowest percent passing in 1941 would by artifact show the most increase to 1956. The correlation between percent passing in 1941 and increase from 1941 to 1956 was found, however, to be only .06.

Table 7 presents a comparison of the data which demonstrates the differential increases made by our subjects in the four classifications of items. Immediate rote memory items and those involving practical reasoning showed smaller increases than those involving abstract reasoning and vocabulary. This finding tends to confirm the conclusion of most investigators that

TABLE 7. Comparison of Four Subscales on Increase from Adolescence to Adulthood

| Scale | Number of Items | Rank order among 26 items | | Mean of Increases in Percent Passing |
|---|---|---|---|---|
| | | *M* | Mdn. | |
| Vocabulary | 8 | 8.1 | 7.3 | 43.7 |
| Reasoning | | | | |
| Abstract | 6 | 9.9 | 11.8 | 42.2 |
| Practical | 7 | 17.4 | 18.0 | 25.1 |
| Memory | 5 | 21.0 | 22.0 | 18.7 |

growth of functions dependent on continued learning and on the building of concepts is more likely to continue into adulthood than is growth of memory or performance functions.

These data make it clear that one should not present conclusions about the growth of intelligence into the adult years without specifying what kinds of intelligence one means and how one is measuring these. If one administered an intelligence test consisting solely of items measuring vocabulary and abstract reasoning, he would find more increase of "intelligence" into adulthood than if he used a test heavily weighted with memory and practical reasoning items. Thus one cannot realistically expect ever to describe the shape of *the* mental growth curve.

A finding in these data which has important implications when it comes to trying to predict amount of subsequent mental growth in an individual adolescent is the negative relationship between level of ability at adolescence and gain in mental age score from adolescence to adulthood. We have found, for example, that the higher the 1941 Stanford-Binet score, whether expressed as IQ or MA, the slightly lower the rate of increase (MA increase/CA increase) to 1956 for the group of 45 subjects who were 14 to 16 years of age in 1941. (For this analysis, subjects less than 14 years of age in 1941 were eliminated in order to minimize the operation of differential rates of mental growth in adolescence.) The correlations expressing this ($r = -.152$ and $-.212$ for IQ and MA, respectively) are not significantly different from zero. Comparable correlations for males and females considered separately are consistent with the findings on sex differences discussed earlier in that the females showed a significant negative relationship, whereas the males did not. The correlations between 1941 IQs and rate of increase to 1956 for 24 males and 21 females were .058 and $-.385$, respectively.

Thus, although the adult scores on intelligence tests can be predicted with some accuracy from the score one obtains as an adolescent (as indicated by the correlations in Table 1), we have not isolated factors which permit us to predict the amount of mental age gain. The fact that adult intelligence can be predicted from adolescent intelligence with such a high degree of accuracy despite the lack of relation between adolescent intelligence and amount of gain to adulthood is, of course, due to the fact that the amount of gain at these levels represents such a relatively small proportion of the total "amount" of intelligence. We can anticipate that there will be an increase in mental age from adolescence to adulthood, but we cannot predict which subjects will gain most. We do know that amount of gain is not dependent upon level of initial mental age, on the extent of education, or on the level of ancestral intelligence (Bradway & Robinson, 1961), in contrast to mental growth in the earlier years where these factors are important (Bradway, 1945a). Our only positive finding in this area is the one pointed out earlier, which is that adolescent girls with high IQs can be expected to increase less in mental age than will girls who test lower, or than males who test at any level.

## Prediction of Adult Intelligence

The next questions which arise are those of how early can adult intelligence be predicted, and what are the best measures for predicting it. Or, what are the origins of adult verbal and nonverbal abilities?

An analysis of the predictive value of Stanford-Binet preschool items to adult WAIS scores can help answer these questions. At the time of the 1941 study, the principal investigator devised four subscales from the preschool Form L and Form M Stanford-Binets and computed four subscale quotients for each subject (Bradway, 1945b). These were labeled "verbal quotient," "nonverbal quotient," "memory quotient," and "number concept quotient." Because the Stanford-Binet at the adolescent and adult levels is an almost wholly verbal instrument, it was not possible in that earlier study to determine whether or not one maintains his relative superiority and inferiority in these categories over any length of time. The administration of the Wechsler Adult Intelligence Scale in the present study, however, allows us to make such "profile" comparisons. Corresponding subscale indices were derived from the WAIS by using Comprehension, Similarities, Information, and Vocabulary for comparison with the 1931 verbal quotient; the Performance scale for comparison with the nonverbal quotient; Digit Span with the memory quotient; and Arithmetic with the number concept quotient. Inasmuch as the WAIS results are in standard scores, all that was necessary to obtain indices for these four WAIS categories which were comparable with each other was to divide the verbal total by four and the performance total by five. The mean of these four indices was then obtained for each subject, and his score for each category was expressed as a deviation from his own mean. A similar method was used for the 1931 scores, except that each of the four subscale scores had to be converted into standard scores before the individual profiles of deviation scores could be figured.

TABLE 8. Correlation of Predictive Preschool Items of Forms L and M of the Stanford-Binet with WAIS Verbal Quotients

| Location | Item | $r_{\text{bis}}$ | Groups by Years of Age[a] | Type[b] |
|---|---|---|---|---|
| M II ½ 6[c] | Obeying Simple Commands | .93°°° | 2½ | V |
| L III 2[c] | Picture Vocabulary | .77°° | 2½ | V |
| L III 4[c] | Picture Memories | .85°°° | 2½ | M |
| M III 3 | Objects by Use | .66° | 2½ | V |
| M III 6 | Repeating Three Digits | .67° | 2½ | M |
| L III ½ 2 | Picture Vocabulary | .84°° | 4 | V |
| L III ½ 5 | Objects by Use | .98°°° | 2½ | V |
| L III ½ 6[c] | Comprehension I | .86°° | 4 | V |
| M III ½ 6 | Comprehension I | .74°° | 2½ | V |
| L IV 1[c] | Picture Vocabulary | .85°°° | 4 | V |
| L IV 2[c] | Objects from Memory | .79°° | 4 | M |
| L IV 4[c] | Pictorial Identification | .80°° | 4 | V |
| L IV A[c] | Memory for Sentences I | .80°° | 4 | M |
| M IV 6 | Memory for Sentences I | .70° | 2½ | M |
| L IV½ 6 | Opposite Analogies I | .73°° | 4 | V |
| M IV½ 1 | Animal Pictures | .64° | 4 | NV |
| M IV½ 5 | Materials | .74°° | 4 | V |
| L V 5 | Memory for Sentences II | .64° | 4 | M |
| M V 2 | Number Concept of Three | .74°° | 4 | NC |
| M V 3[c] | Pictorial Similarities and Differences | .61° | 5½ | V |
| L VI 4 | Number Concepts | .71°° | 5½ | NC |
| L VI 5 | Pictorial Likenesses and Differences | .72°°° | 5½ | V |
| M VI 2 | Copying Bead Chain | .55° | 5½ | NV |
| M VI 3[c] | Differences | .60° | 5½ | V |
| M VII 4[c] | Three Digits Reversed | .62° | 5½ | M |

[a] There were 17 subjects in the 2½-year-old group; 16 in the 4-year-old; 17 in the 4½-year-old; and 21 in the 5½-year-old. The other groups (2-year-olds, 3-year-olds, 3½-year-olds, and 5-year-olds) were not large enough for valid calculations.

[b] These refer to the Bradway subscales described previously (Bradway, 1945b).

[c] Indicates used in 1960 L-M Revision of the Stanford-Binet.

° Significant beyond .05 level.

°° Significant beyond .02 level.

°°° Significant beyond .01 level.

The stability of pattern over a period of 25 years was tested by the correlation between the deviation scores calculated as each individual's deviation from his own mean for the 1931 subscales, and the deviation scores also calculated as each individual's deviation from his own mean for the 1956 WAIS. Three of these four correlations are significant in the positive direction. For the verbal component, the Pearsonian *r* is .236 significant at the .02 level; for the nonverbal component, .191 significant at the .05 level; and for the memory component, .301 significant at the .01 level. For the number concept component, (figured on 79 subjects rather than 110 because for 31 subjects no such ratio could be computed at the preschool ages) the correlation was only .089, not significant, suggesting that relative strength or weakness in this ability is not stable over a 25-year period. Inasmuch, however, as one's score on this scale contributes to his deviation scores on the other three, further analysis is necessary before drawing such a conclusion.

To study the predictive value of preschool items in more detail, biserial correlations of each Stanford-Binet preschool item to adult WAIS verbal quotient and adult WAIS performance quotient were calculated. This anal-

TABLE 9. Correlations of Predictive Preschool Items of Forms L and M of the Stanford-Binet with WAIS Performance Quotients

| Location | Item | $r_{bis}$ | Groups by Years of Age[a] | Type[b] |
|---|---|---|---|---|
| M III 2[c] | Picture Vocabulary | .90°°° | 2½ | V |
| L III½ 6[c] | Comprehension | .75° | 4 | V |
| L IV 1[c] | Picture Vocabulary | .75° | 4 | V |
| L IV 2[c] | Objects from Memory | .71° | 4 | M |
| L IV 4[c] | Pictorial Identification | .87°°° | 4 | V |
| L IV 5[c] | Discrimination of Forms | .71° | 4 | NV |
| M IV 3[c] | Opposite Analogies I | .70° | 4 | V |
| M IV 4[c] | Pictorial Identification | .70° | 2½ | V |
| M IV 6 | Memory for Sentences I | .75° | 4 | M |
| L IV½ 4 | Materials | .68° | 4½ | V |
| L IV½ A | Pictorial Identification | .71° | 4 | V |
| M IV½ 1 | Animal Pictures | .63° | 4 | NV |
| M IV½ 3 | Repeating Four Digits | .70° | 2½ | M |
| M IV½ 5[c] | Materials | .69° | 4 | V |
| L V 4[c] | Copying a Square | .74° | 4 | NV |
| L V 6 | Counting Four Objects | —.62° | 4½ | NC |
| M V 2 | Number Concept of Three | .75°° | 4 | NC |
| M V 5 | Comprehension II | .78°°° | 5½ | V |
| L VI 1 | Vocabulary | .86°°° | 5½ | V |
| L VI 2 | Bead Chain from Memory | .68° | 4½ | M |
| L VI 3[c] | Mutilated Pictures | .58° | 5½ | |
| | | .82°° | 4½ | NV |
| | | .77°°° | 5½ | |
| L VI 5 | Pictorial Likenesses and Differences | .53° | 5½ | V |
| M VI 2 | Copying Bead Chain | .94°°° | 5½ | NV |
| M VI 5 | Counting 13 Pennies | .67°° | 5½ | NC |
| M VI 6 | Opposite Analogies I | .73°° | 5½ | V |
| L VII 1[c] | Picture Absurdities I | .61° | 5½ | V |
| M VII 4 | Three Digits Reversed | .60° | 5½ | M |

[a] There were 17 subjects in the 2½-year-old group, 16 in the 4-year-old, 17 in the 4½-year-old, and 21 in the 5½-year-old. The other groups (2-year-olds, 3-year-olds, 3½-year-olds, and 5-year-olds) were not large enough for valid calculations.

[b] These refer to the Bradway subscales described previously (Bradway, 1945b).

[c] Indicates used in 1960 L-M Revision of the Stanford-Binet.

° Significant beyond .05 level.

°° Significant beyond .02 level.

°°° Significant beyond .01 level.

ysis is summarized in Tables 8 and 9 in which are listed those items whose biserial *r*'s were significant beyond the .05 level for the indicated age groups.

The figures in the column headed "Groups by years of age" indicate the age group for which the corresponding biserial *r* was figured. Since there was no way of combining the age groups, biserial *r*'s were figured for each 1931 age group for which an item was pertinent, which was defined for purposes of this analysis as each age group in which the number of subjects was 15 or more and for which the percent passing the item was between 25 and 75. The distribution by type of the total 79 items for which biserial *r*'s at one or more age levels were computed were as follows: 37 verbal, 24 nonverbal, 12 memory, and 6 number concept. The type designation follows that previously described by Bradway (1945b). Comparison of the frequency of each type found to have significant predictive value at one or more age levels and the total number of such items studied shows the following: of 37 verbal items, 14 (38%) predicted WAIS verbal IQ and 14 (38%) predicted WAIS performance IQ; of the 28 nonverbal items, 2 (8%) predicted WAIS verbal IQ and 5 (21%) predicted WAIS performance IQ; of the 12 memory items, 7 (59%) predicted WAIS verbal IQ and 5 (41%) predicted WAIS performance IQ; of the 6 number concept items, 2 (33%) predicted WAIS verbal IQ and 3 (50%) predicted WAIS performance IQ. This kind of analysis is only suggestive because the number of items is small, but it supplements in more detail the findings for the subscales reported above. In summary: verbal items predict both verbal and performance adult ability better than do nonverbal items; each kind of memory item (memory for digits forward and reversed, for sentences, for objects, for pictures, and for bead chain) predicts both WAIS verbal and performance ability; number concept items are inconsistent in their predictive value—the only significant negative correlation obtained being for a number concept item.

When one compares these results with those for the subscale deviation scores reported above, similarities are found. Both analyses suggest the greater value of verbal and memory items over nonverbal items in predicting adult abilities, and the unreliability of number concept items at the preschool level as predictors of adult abilities.

## Measurement of Intelligence

After reviewing these data we are in a better position to approach the problems of what index or indices to use in expressing adult intelligence and how best to predict it. If the 1937 Stanford-Binet is used to measure adult intelligence, a use for which it was not specifically designed, it is apparent that a correctional factor of some kind must be applied to IQs obtained according to the published instructions. Merrill kindly applied the deviation scores provided in the 1960 Revision (Terman & Merrill, 1960) to our data.

The resulting intelligence quotients are 109.3 in 1931, 109.4 in 1941, and 119.4 in 1956. These findings of identity between preschool and adolescence and a 10 point rise to adulthood indicate that the correction inherent in the 1960 deviation scores does not solve the problem. For our research purposes we rejected the use of a multiplicative factor because this would have changed the standard deviations, and used instead a subtractive factor which, since the standard deviations were nearly identical, was sufficient to equate the adult set of scores with the two previous ones. This, although a satisfactory solution for research purposes, may not be the optimal solution from the point of view of the clinician. Perhaps the possibility of revised deviation tables, accounting for greater increase into adulthood than do those published with the 1960 Revision, should be explored.

This problem is not encountered in measuring adult intelligence with the WAIS which was specifically designed for the measurement of adult intelligence. Moreover, our data demonstrate that there are different kinds of ability which have a degree of constancy throughout life. The structure of the WAIS is a recognition of this.

When new tests are constructed, those for adults would do well to provide valid measurements of different kinds of abilities, because the practicing clinician needs such information to evaluate selective impairment of intellectual functioning and for vocational guidance and occupational placement. In contrast, a preschool test heavily weighted with verbal items should be a more valid predictor of adult abilities, both verbal and nonverbal, than one utilizing many performance-type items. Moreover, the former would be consistent with the usual purpose for which preschool tests are used in that the prediction of school performance puts a heavy demand on verbal abilities.

The present data demonstrate that the defining of adult intelligence is complex. It is not expressed by a single index in terms of abilities, in terms of one's place relative to others of the same age group, or of one's location on a scale of development and decline. As Cameron and Magaret (1951, p. 158) have said in alluding to the use of tests in defining behavior, "Performance on test items alone can no more define biosocial immaturity than can a refractive index define the complex process of seeing."

## Summary

A group of 111 subjects who had been tested with the Stanford-Binet (S-B) at the preschool and adolescent levels were administered the S-B and WAIS at adulthood (25 years after initial testing). Correlations of preschool IQs with adult S-B and full WAIS IQs are .59 and .64. Correlations of

adolescent IQs with adult S-B and full WAIS IQs are .85 and .80. The S-B IQ shows a mean increase of 11 points from adolescence to adulthood, which corresponds to a mental age gain of 20 months beyond what would be expected if mental growth ceases at 16 years. The data were analyzed in an attempt to determine the point at which mental growth, as measured by these tests, ceases, but it could not be determined.

The males show significantly more IQ gain from adolescence to adulthood than do the females ($p < .01$). Factors associated with amount of gain were not isolated except that girls with high IQs at adolescence increase less than do girls who initially test lower or than males who test at any level. There is no evidence that educational attainment is related to amount of gain.

Analysis of increases in percent passing S-B items from adolescence to adulthood indicates that intellectual growth after adolescence is greater in abstract reasoning and vocabulary than in rote memory and practical reasoning.

Analysis of the value of S-B preschool items in predicting adult WAIS scores shows some stability of pattern of individual differences in verbal versus nonverbal versus memory superiority over the 25 years. It also shows the greater value of verbal and memory items over nonverbal items in predicting both verbal and performance adult IQs, and the unreliability of number concept items as predictors of adult abilities.

The data are interpreted as demonstrating that adult intelligence is complex and cannot be meaningfully expressed by a single index.

## REFERENCES

Bayley, Nancy. On the growth of intelligence. *Amer. Psychologist,* 1955, **10**, 805–818.

Bayley, Nancy. Data on the growth of intelligence between 16 and 21 years as measured by the Wechsler-Bellevue scale. *J. genet. Psychol.,* 1957, **90**, 3–15.

Bayley, Nancy, & Oden, Melita H. The maintenance of intellectual ability in gifted adults. *J. Gerontol.,* 1955, **10**, 91–107.

Bradway, Katherine P. IQ constancy on the Revised Stanford-Binet from the preschool to the junior high school level. *J. genet. Psychol.,* 1944, **65**, 197–217.

Bradway, Katherine P. An experimental study of factors associated with Stanford-Binet IQ changes from the preschool to the junior high school. *J. genet. Psychol.,* 1945, **66**, 107–128. (a)

Bradway, Katherine P. Predictive value of Stanford-Binet preschool items. *J. educ. Psychol.,* 1945, **36**, 1–16. (b)

Bradway, Katherine P., & Robinson, Nancy M. Significant IQ changes in twenty-five years: A follow-up. *J. educ. Psychol.,* 1961, **52**, 74–79.

Bradway, Katherine P., Thompson, Clare W., & Cravens, R. B. Preschool IQs after twenty-five years. *J. educ. Psychol.*, 1958, 49, 278–281.

Cameron, N., & Magaret, Ann. *Behavior pathology.* Boston: Houghton Mifflin, 1951.

Davis, F. B. The derivation of three subscores from the 1937 Revision of the Stanford-Binet Scales. *J. consult. Psychol.*, 1941, 5, 287–291.

Inhelder, B., & Piaget, J. *The growth of logical thinking from childhood to adolescence.* New York: Basic Books, 1958.

Miner, J. B. *Intelligence in the United States.* New York: Springer, 1957.

Owens, W. A. Age and mental abilities: A longitudinal study. *Genet. Psychol. Monogr.*, 1953, 48, 3–54.

Terman, L. M., & Merrill, Maud A. *Stanford-Binet Intelligence Scale.* Boston: Houghton Mifflin, 1960.

# 6

# Anxiety, Intelligence, and Achievement in Children of Low, Average, and High Intelligence

## —*John F. Feldhusen and Herbert J. Klausmeier*

*The previous articles have been chiefly concerned with similarities and differences in the intellectual abilities of children and adults. Cognitive characteristics alone, however, cannot account for all of the variation in rate of learning, in interest shown toward classroom activities, or in level of performance. In this paper the authors report an investigation, conducted among fifth-grade children, indicating a relationship between anxiety and both IQ and achievement. Several explanations for the links discovered in that relationship are offered; apparent, however, is the need for further research in order to clarify the influence of anxiety on the development of intelligence, personality, and learning.*

THE RELATIONS among various cognitive abilities and between cognitive and psychomotor abilities of school children are being clarified through widespread and continuous research. Much less is known about the relation between affective characteristics and cognitive abilities. However, anxiety with moderate to high tension, whether artificially induced through experimentally controlled stress or present in the subjects from unspecified sources, has been shown to reduce learning efficiency (*1*, *3*, *5*, *6*). The pur-

Reprinted from *Child Development*, 1962, 33, 403–409, by permission of the senior author and the Society for Research in Child Development, Inc.

pose of the present study was to ascertain the relation between anxiety as measured by a test and IQ and between anxiety and school achievement in children of low, average, and high IQ. Basically, the present study attempts to extend and clarify the work of Sarason and of McCandless, Castaneda, and Palermo with children of elementary school age.

Sarason (8) reported that anxiety correlated negatively with IQ and with achievement. Eight correlations were reported between Otis IQs and scores on the Sarason Test Anxiety Scale ( *7*); the children were enrolled in grades 2 through 5 in two schools. The correlations ranged from −.012 to −.284; four were significant at the 1 per cent level. Of eight correlations, ranging from −.002 to −.294, between scores on standardized achievement tests and the anxiety scores for the same eight groups of children, one was significant at the 5 per cent level and four at the 1 per cent level.

McCandless and Castaneda (*4*) found anxiety related to IQ in sixth grade girls but not in sixth grade boys. Two correlations were reported between Otis IQs and scores on their Children's Manifest Anxiety Scale (CMAS) (*1*). Both correlations were negative, for boys, −.16 and for girls, −.43, the latter being significant at the 1 per cent level. They also reported correlations between anxiety scores and five achievement measures from the Iowa Every Pupil Test for six groups of boys and girls in grades 4, 5, and 6. Of the 30 correlations, 28 were negative, two being significant at the 5 per cent level and 11 at the 1 per cent level. The correlations ranged from .00 to −.74.

The conception of anxiety as a research variable held by Sarason and his associates is psychoanalytic in nature (8). Thus, anxiety is a danger signal of crucial significance in the organism's self-system and in the defense of self. Castaneda, McCandless, and Palermo (2) assume instead that the anxiety scale provides a method of measuring level of drive or motivation.

The present researchers treat anxiety as a generalized and diffused fear response to many aspects of the environment. Thus, the organism perceives real or imagined phenomena as threatening to the self although the non-anxious does not so perceive them. Sustained and unresolved fears, not connected with specific objects or events, may evolve genetically through real and threatening experiences with the physical and social environment. The usual definition of fear connotes a real and present threatening object, situation, or person. In anxiety, the fear response becomes dissociated from any real and objectively determinable threat to the organism. In the absence of such a threat, a resolution or culmination by overcoming or retreating from the threat is obviously impossible. No attempt is made here to explain the dissociation process except to suggest that dissociation may result from the organism's failure to overcome the threat, inability to retreat without ego danger or threat to the self-concept, or an overwhelming persistence of a set of fear-evoking stimuli. Dissociation may, thus, be the only defense mechanism left to the organism, but it fails to relieve the essential fear which becomes general and diffuse.

## Subjects

Eighty average and high IQ children were drawn from approximately 450 children enrolled in 15 regular fifth grade classes in nine public schools. Forty low IQ children were drawn from approximately 224 children enrolled in 14 special classes for mentally retarded educable children. The main criteria of selection were IQs, as ascertained with the Wechsler Intelligence Scale for Children, and birthdate within a 15-month interval. Thus, the sample included 20 boys and 20 girls with WISC IQs of 56 to 81, 20 boys and 20 girls with WISC IQs of 90 to 110, and 20 boys and 20 girls with IQs of 120 to 146. The IQ tests were administered by school or clinical psychologists during a four-month interval which coincided with the collection of the other data reported in this study. Low IQ children were excluded who exhibited the characteristics of definite organic types such as mongoloid or cretin; who were handicapped orthopedically or in hearing, vision, or speech; or who were suffering from epilepsy.

## Procedure

The CMAS as reported by Castaneda, McCandless, and Palermo (2) was administered by the authors and several graduate students. In the regular classrooms, the examiner read the directions aloud and then the children completed the scale. In testing the mentally retarded, the examiner read the directions and then each item of the scale, successively, pausing to give the children, in small groups or as individuals, time to respond to the item or to give the teacher or other assistants time to explain a word or phrase of the item.

Reading, arithmetic, and language achievement were measured through use of the California Achievement Battery, 1950, Form AA. The Primary Form was used with the low IQ children, the Elementary Form, with the average and high IQ children. The general procedure was to administer the entire battery to the entire classroom group of which the subjects were a part. The tests of the low IQ children were, however, administered individually or in small groups utilizing special supplementary instructions developed by the authors. All of the achievement testing was done by the authors and graduate assistants.

## Results

Table 1 shows the mean anxiety level for children of low, average, and high intelligence to be 20.2, 14.8, and 12.0, respectively. An analysis of variance, as reported in Table 2, shows that the differences among the three IQ group means are significant at the 1 per cent level. A subsequent Tukey gap test, as reported in Table 3, shows that the difference between the low IQ and the high IQ groups in mean anxiety, 8.2, is significant at the 1 per cent level, the difference between the low and the average, 5.4, significant at the 5 per cent level, and the difference between the average and the high, 2.8,

TABLE 1. Means for Anxiety in Children of Low, Average, and High Intelligence at the Fifth Grade Level

| Group | *N* | Mean |
|---|---|---|
| IQ 56 to 81 | 40 | 20.2 |
| IQ 90 to 110 | 40 | 14.8 |
| IQ 120 to 146 | 40 | 12.0 |
| Boys | 60 | 13.8 |
| Girls | 60 | 17.5 |

TABLE 2. Analysis of Variance for Anxiety in Fifth Grade Children of Low, Average, and High Intelligence

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Sex | 1 | 396.0 | 5.89* |
| IQ groups | 2 | 688.7 | 10.25** |
| Sex by IQ | 2 | 3.3 | .05 |
| Error | 114 | 67.2 | |

* Significant at the 5 per cent level.
** Significant at the 1 per cent level.

TABLE 3. Analysis of Differences in Mean Anxiety Level Among the Three IQ Groups Using a Tukey Gap Test

| Group | Mean | Mean − 12.0 | Mean − 14.8 |
|---|---|---|---|
| Low IQ | 20.2 | 8.2** | 5.4* |
| Average IQ | 14.8 | 2.8 | |
| High IQ | 12.0 | | |

* Significant at the 5 per cent level.
** Significant at the 1 per cent level.

is not significant. In Table 1, it is also seen that the mean anxiety level for girls, 17.5, exceeds that for boys, 13.8. The difference is significant at the 5 per cent level as shown in Table 2.

A further analysis of the relations of anxiety to IQ within each IQ group is reported in Table 4. The three correlations for the low, average, and high IQ groups are −.28, −.35, and −.07, respectively. The only correlation significant at the 5 per cent level is the −.35 for the average IQ group.

TABLE 4. Correlations of Anxiety with Intelligence and Achievement Scores for Fifth Grade Children

| | | Correlation of Anxiety with | | | |
|---|---|---|---|---|---|
| Group | *N* | WISC IQ | Reading | Arithmetic | Language |
| IQ 56 to 81 | 40 | −.28 | −.13 | −.49°° | −.27 |
| IQ 90 to 110 | 40 | −.35° | −.31° | −.20 | −.33° |
| IQ 120 to 146 | 40 | −.07 | +.15 | −.02 | +.12 |
| Boys | 60 | −.47°° | −.48°° | −.47°° | −.49°° |
| Girls | 60 | −.40°° | −.38°° | −.40°° | −.43°° |

° Significant at the 5 per cent level.
°° Significant at the 1 per cent level.

Table 4 also presents the results of the analysis of the relation between anxiety and achievement in reading, arithmetic, and language for the three IQ groups. Seven of the nine correlations are negative. The correlations of anxiety with reading and language in the average IQ group, −.31 and −.33, respectively, are significant at the 5 per cent level, and the correlation of anxiety with arithmetic in the low IQ group, −.49, is significant at the 1 per cent level.

Finally, Table 4 shows eight correlations of anxiety with IQ and achievement scores for the 60 boys and 60 girls in this study. The eight correlations are negative ranging from −.38 to −.49, all significant at or beyond the 1 per cent level. Little difference is shown between boys and girls except that the correlations for girls, −.40, −.38, −.40, and −.43, are consistently lower than the corresponding correlations for boys which are −.47, −.48, −.47, and − .49. It should be noted that the distributions of IQ and achievement scores by sex were trimodal because of the initial method of selection of subjects by IQ, and this condition tends to inflate these correlations for the sex groups.

## Discussion

The results of the present study are similar to those obtained by Sarason (*8*) and by McCandless and Castaneda (*4*) in that negative correlations were obtained between anxiety and cognitive abilities. However, the present study revealed also (a) a significant negative correlation between anxiety and IQ in both boys and girls whereas McCandless and Castaneda found it only in girls; (b) significantly higher mean anxiety scores for girls than for boys; (c) the same pattern and about the same size of correlations of anxiety with IQ and three achievement measures for both sexes; and (d) significantly higher mean anxiety scores for low IQ children than for the average and high IQ with the difference between the latter two groups not being significant. Thus, the relation of anxiety to cognitive abilities in three clearly demarcated IQ groups and according to sex is clarified in the present study.

Superior mental ability may make it possible for a child to assess more adequately the real and present danger in any current threatening object, situation, or person. Thus, his fears may be specific and ascertainable and unrelated to variations in an already high mental ability. Children of low and average IQ may show greater anxiety or generalized fear and a close relation between IQ and anxiety because of their limited intellectual capacity to evaluate the threat or danger in objects, situations, or people that confront them. In this study, an upper limit of intellectual advantage seemed to be reached at an IQ of 120, for a nonsignificant correlation was found with IQs of 120 to 146.

Possibly a low or average child who is more anxious is less free to respond to the teacher, the learning material, and the learning situation than is a child who is less anxious. The energies of the anxious child may be expended partially in the general fear responses to various aspects of the school environment. Thus, rather than providing drive or motivation to learn, anxiety is conceived to be debilitating, an interference with the learning process.

A relatively high negative correlation, −.49, was shown between anxiety and arithmetic achievement in the low IQ children enrolled in special classes for the mentally retarded. This suggests that the teachers of these children should be cautious not to produce anxiety in the lowest achievers who are already highly anxious. If anxiety is debilitating, subsequent arithmetic achievement of these children will become more depressed.

The level of anxiety may also be an indicator of the successes which the child experiences. If anxiety is greater in children whose achievements in reading, arithmetic, and language are lower, then lack of success may be associated with the generalized fear response. Failure and fear may thus be unavoidable conditions confronting children in the low and average IQ groups for a substantial part of their daily lives. Little is known about the possible effects of such a psychological regimen. Further research is needed urgently to clarify the role of anxiety in the development of intelligence, personality, and learning.

## Summary

Forty children with WISC IQs of 56 to 81, 40 children with WISC IQs of 90 to 100, and 40 children with WISC IQs of 120 to 146, divided equally by sex in each group, were used as subjects in an analysis of the relationship of anxiety as measured by the Children's Manifest Anxiety Scale to IQ and achievement. Significantly greater mean anxiety was found in the low IQ group than in the average or high, the difference between the latter two groups not being significant. The correlation between anxiety and IQ was significant in the average IQ group, nearly significant in the low IQ group. Significant correlations were found between anxiety and reading achievement in the average IQ group, between anxiety and arithmetic achievement in the low IQ group, and between anxiety and language achievement in the average IQ group. The six correlations between anxiety and achievement measures for the low and average IQ children were negative. Two of the three correlations between anxiety and achievement for the high IQ children were low and positive, the third was very small and negative. The eight correlations of anxiety with IQ and achievement scores for boys and girls were negative and significant at the 1 per cent level, but the mean anxiety level for girls was significantly higher than for boys.

## REFERENCES

1. Ainsworth, L. H. Rigidity, insecurity, and stress. *J. abnorm. soc. Psychol.*, 1958, **56**, 67–74.
2. Castaneda, A., McCandless, B. R., & Palermo, D. S. The children's form of the manifest anxiety scale. *Child Develpm.*, 1956, **27**, 317–326.
3. Chansky, N. M. Threat, anxiety, and reading behavior. *J. educ. Res.*, 1958, **51**, 333–340.
4. McCandless, B. R., & Castaneda, A. Anxiety in children, school achievement, and intelligence. *Child Develpm.*, 1956, **27**, 379–382.
5. Palermo, D. S., Castaneda, A., & McCandless, B. R. The relationship of anxiety in children to performance in a complex learning task. *Child Develpm.*, 1956, **27**, 333–337.
6. Sarason, I. G. Test anxiety, general anxiety, and intellectual performance. *J. counsel. Psychol.*, 1957, **21**, 485–490.
7. Sarason, S. B., Davidson, K., Lighthall, F., and Waite, R. R. A test anxiety scale for children. *Child Develpm.*, 1958, **29**, 105–113.
8. Sarason, S. B., Davidson, K. S., Lighthall, F. F., Waite, R. R., and Ruebush, B. K. *Anxiety in elementary school children.* New York: John Wiley, 1960.

# 7

# Exploring Children's Thinking — *Leonore Boehm*

*According to the Swiss psychologist Jean Piaget, the young school child is characterized by egocentric thinking; that is, he attributes to objects and to other people the same experiences, perceptions, actions, motivations, and thoughts that he has. "He believes that the world appears the same to everybody and that it is and remains what it appears to him at each particular moment." Drawing upon the rich theoretical and empirical work of Piaget, Leonore Boehm discusses in the present article characteristics of children's thinking in both childhood and adolescence. The implications she draws in this area are of primary interest to educators; and involve curriculum planning, the ways in which children see themselves, and the world and the ways in which they think.*

FOR MORE than thirty years child psychologists at the Institute of Educational Sciences in Geneva, Switzerland, have been studying children's thinking. Many of the findings, especially those of Jean Piaget and Barbel Inhelder, are important to the elementary-school educator because of the meaning they have for curriculum planning and for methods of teaching.

To learn how children's thinking develops, Piaget and his staff have been interviewing boys and girls who attend the Children's School at the institute and children enrolled at other private or public schools.

Piaget has described not only how the child thinks at various ages, but also why he thinks as he does. With his special method of investigation, Piaget reveals thought processes hidden behind the child's answers. While

Reprinted from *Elementary School Journal*, 1961, **61**, 363–373, by permission of The University of Chicago Press. © The University of Chicago Press.

most methods of investigation disclose only the outcome of the child's reasoning, Piaget, through his method, has gained insight into the causes of children's reasoning, into mental mechanisms.

The investigators interview each child individually; and, once the interview is launched, each question they ask is based on the child's response to the preceding question. The investigators probe, asking many questions about the same response. Thus there cannot be complete uniformity in the questions asked or in the number of responses for each question. Unfortunately, some writers have not understood this method or its purpose and advantage.

### The Child and the World . . . One

According to Piaget, the young school child is still egocentric. Piaget does not use this term in the customary sense. To him it means that the child is not yet able to distinguish fully between himself and the rest of the world—people or things.

The small child who does not yet differentiate between the self and the non-self cannot be objective. He believes that the world appears the same to everybody and that it is and remains what it appears to him at each particular moment. To him our fragmentary and temporary perceptions are absolute, rather than dependent on time, space, and experience. He is not aware that a mountain seen from one position looks different when viewed from another position. Rather he believes that while he moved, the mountain changed its shape.

Because the child does not differentiate between people and things, people remain the prototype of objects, which he perceives as having human characteristics and as being alive.

### Why Do Bridges Break?

He endows the world with intentions and a sense of morals. In his view of the world there is no fortuity: he believes that all events have a reason, a purpose, and that there are no chance occurrences. A bridge that breaks while the child is crossing does so to punish the child, who was naughty. To him everything can be explained, everything is motivated.

In the child's eyes, the world, or reality, then, is alive ("animism"), and it has the same magical powers and qualities that adults have. The moon follows the child purposefully. The wind is alive. How else could it blow? Clouds know that they move. They, too, accompany the child.

To the child, adults know all and are all-powerful. The world, to him, is man-made ("artificialism"), and its maker may be the child's father, grandfather, or great-grandfather.

In the child's view, inner experiences, feelings, thoughts, dreams, are undistinguishable from outer experiences and have the same properties that objects have ("realism"). A child who wants strawberries for supper when there are none in the house wakes up happy and satisfied the next morning: his wish was fulfilled during the night. He ate strawberries. What to us is a dream, to him is reality. There is, then, in the very young child complete confusion between the psychical and the physical order of things—confusion between psychological intentions, wishes, and desires, and physical reality.

In the child's world everything must be purposeful; rather than look for true explanations the child tries to justify all happenings. Because he does not differentiate between himself and the world, he cannot search for actual temporal, causal, or logical relations but establishes cause and effect from chance occurrences.

## One Thought . . . And Another . . .

The child's thoughts follow one another without connection or relation ("juxtaposition"). Often he thinks of two factors alternately rather than simultaneously. Pour liquid from one container into another of a different shape, and ask the child whether the amount of liquid has changed. He will tell you that it has. Or tip the container so that the level of the liquid changes, and again ask the child whether the amount of the liquid in the container has changed. Again, he will tell you that it has.

Because he focuses his attention on one aspect, he often makes contradictory responses successively. At one time he compares only the heights of the two containers and decides that there is less liquid in the shorter one. Soon afterward he compares the widths, forgetting the heights, and decides that there is more liquid in the wider container. Thus he misjudges the amount of liquid both times.

The child cannot distinguish related parts from the whole; he sees them as independent of one another and of the whole. For example: if the child is told that all apples and pears are fruit, he believes that all fruit are either apples or pears. The child whose first furry, four-legged animal is called *bow-wow* by his elders is likely to think that all other furry, four-legged animals are bow-wows and so calls them until his elders teach him to distinguish between furry, four-legged animals that are bow-wows and furry, four-legged animals that are meows or horsies or moo-moos. An object placed at the child's right cannot in his thinking at the same time be at somebody else's left. A cow cannot be an animal and a mammal. Egocentric thought, then, is subjective, intuitive, impressionistic, non-analytic, and absolute.

## Yesterday's Child . . . And Today's

In studying children's thought thirty years ago Piaget found that at the age of eight there was a decrease in egocentric thinking in regard to concrete materials or problems the child could solve by handling objects. Today, Piaget and his staff believe that modern children may outgrow egocentric thought earlier than in the past because they are exposed to many more experiences through better methods of teaching and through such media as TV and radio.

Up to the age of ten the child's reasoning is still linked to action. The child needs to look at objects. He needs to handle them. The younger the child, the more he learns through his senses. The infant learns to distinguish one thing from another by tasting, by touching, by smelling. The older child learns to distinguish by handling objects. Problems that he can solve by experimenting, he may not be able to solve on the verbal—that is, the abstract—level of thought. At ten and eleven the child thinks logically about fairly simple concrete matters even without experimentation.

A child younger than seven cannot serialize rods of different length. Give a child the following problem: "Edith is fairer than Susan; Edith is darker than Lily. Who is the darkest of the three?" The child under twelve will not be able to answer correctly, for the question requires serialization on a formal level.

## Child and World . . . No Longer One

Gradually the child learns to distinguish between himself and the world, between his ideas and those of others. He gains insight into the perspective of others and realizes that his perceptions and evaluations are subjective. He "depersonalizes" the world and recognizes that many events are fortuitous. According to Piaget, reality is now "dissubstantialized."

Thus the child no longer needs to justify or explain all occurrences and turns instead to a search for true cause-and-effect relationships. He realizes that the bridge did not break down to punish the naughty child: the bridge broke down because the wood had rotted. He distinguishes accidental occurrences from cause-and-effect happenings and knows that many events are unrelated to his behavior. It might rain on the day of a picnic for reasons outside himself, rather than because of his mischievous doings. He distinguishes right from wrong and accidental mishaps from intentional ones, being now more concerned with the motivation behind an action than with its consequences.

Thus he no longer thinks in absolutes. He realizes that the world is a world of relations, and he recognizes interrelatedness where it exists ("dissubjectivation of causality"). He can form concepts about the world. He sees general principles and establishes an integrated system of thought, a coherent reference system.

## Apples . . . Pears . . . Fruit

Because the child can now find common elements in given factors, he can classify; he can distinguish between higher and lower elements of given factors and build hierarchies. He understands that, while apples and pears are fruit, all fruit are not apples and pears. He realizes that while apples and pears can be eaten, not all fruit can be.

He can differentiate ordinal and cardinal classes and establish series, relating corresponding elements.

His thinking has become logical because he has learned to check his hypotheses by reversing his thought processes. To use again the example of the containers, the child can now mentally pour the liquid from the second container back into the first and come to the realization that in spite of the appearance the amount of liquid is unchanged. Nothing has been added, nothing has been removed. This reasoning has become an "operation," which Piaget defines as an internalized mental action. To the reality that he perceives with his eyes, the child adds elements of mental construction that cannot be perceived because they are not concrete and thus are elements of abstract thinking.

When the young child watches sugar dissolve in water, he assumes that the sugar becomes extinct because he can no longer see it. Later he comes to realize that the sugar could not have disappeared; it exists in particles too small to be seen. When a kernel of corn pops in a burner, he understands that it increases in size only; in spite of appearances it cannot increase in corpuscular volume.

Throughout adolescence this type of thinking—that is, abstract thinking—develops further. Piaget calls it "formal operation." The adolescent's thinking becomes more and more independent of the concrete world, of his perceptions, and instead of reasoning about concrete things, he forms "mental constructs." Because the formal operation is completely mobile and reversible, all assumptions can be tested mentally. The system of thought of the older adolescent is a coherent structure, an integrated system. He no longer depends on his past experience. Now he can anticipate events and their consequences: he can use foresight. He enjoys meditating on theories and postulates.

### A Principle in Physics

Let us examine a few examples of development in the power to reason. Our examples will be drawn largely from the field of physics, especially the concept of "invariance" or "invariability"—the idea that a substance or an amount to which nothing has been added and from which nothing has been removed remains unchanged in weight or quantity.

The concept is not part of the thinking of the young child. As far as he is concerned, if the appearance of an object changes, if the shape changes, quantity and weight also change.

As long as the child focuses on the most obvious phenomenon—be it the height or the width of the vessels mentioned earlier—as long as the child can think of only one aspect at a time, he does not understand the principle of the preservation of matter, the "invariability" of substance. In the example we are using, he must learn to consider both the height and the width of the containers and their relationship to the amount of liquid in them. He must free himself from his illusory perceptions and take less conspicuous factors into consideration.

### Steps Toward Understanding

Not all features of invariability are grasped at the same age. At seven or eight years of age the child knows that when nothing has been added or taken away, substance and quantity remain the same regardless of appearance. At eight or nine the child is aware of the constancy or duration of space and time in their simple aspects. At nine or ten he knows that, if nothing has been added or taken away, weight remains unchanged, in spite of a change in appearance, and at eleven or twelve he understands the constancy of volume.

### Sizes and Shapes

Give a child younger than seven or eight years of age two containers of the same size and shape. Let him take one bead in each hand and drop the beads into the two containers at the same time. Let him continue in this way till all his beads are in the containers. Then ask the child to transfer the contents of one container into a third container that is different in shape from the other two. Noting the difference in shape, the child believes that the quantity of beads has changed, too. He focuses on only one aspect of dimension, the level of the beads in the third container, and his perception tricks him. He does not realize that the number of beads remains unchanged. He does not know that when nothing is added or subtracted a quantity remains unchanged. It does not occur to him to test his thinking or perception by reversing his procedure—by actually putting the beads back into the first container or by carrying out this reversal mentally.

Up to the age of seven the child does not perceive size as constant. If the child places two rods of the same length side by side and then lets one protrude, he does not understand that the rods remain equal in length, nor does he understand why. The concept of seriation of sizes is achieved at the age of seven or eight. Children of this age can arrange a number of rods of different sizes according to length.

Place before a child younger than nine years of age two lumps of clay that are equal in weight. While the child watches, change the shape of one of the lumps, and the child will think that the lumps no longer weigh the same. He will explain that the weight of the lump has changed because the second lump is longer or shorter.

Ask the younger children whether the substance of the two lumps has changed, and they will tell you that it has, again pointing out as evidence that one lump is shorter or longer than the other.

Place two lumps of clay in water, and change the shape of one lump. The child under eleven or twelve years of age believes that the volume has changed. Ask him why, and he will tell you "because the second lump is shorter (or longer) than the first one."

The child understands the concept of invariability or constancy or permanency of matter only when he realizes that he can reverse the procedure of the experiment to reproduce the original quantity, shape, weight, or volume.

## Parts and Wholes

Because the young child thinks in absolutes and focuses his attention on one factor, he cannot consider the relationships of two classes or the relationship of the subclass to the whole class or of the parts to the whole. He thinks of parts as independent of one another and of the whole. When he considers the whole, he cannot distinguish its parts.

One problem Piaget gave children will serve as an illustration. Piaget placed many brown beads and a few white beads before a child. The child recognized that all the beads were wooden. Piaget then separated the beads into two unequal groups according to their color and asked the child, "Which will make a longer necklace—the wooden beads or the brown beads?" Up to the age of seven the children who were asked the question insisted repeatedly, "The brown ones." They made this response even when they were reminded that all the beads were wooden.

This is an example of perceptive illusion: the color of the beads is more obvious to the children than the substance of the beads. In Piaget's terms, the children "center" their attention on one factor alone, in this case, color.

Also, as we have seen, children cannot form inclusions. They cannot think: brown wooden beads plus white wooden beads equals wooden necklace. They do not yet heed the part-whole relationships but consider only the parts.

The seven-year-old Genevan child knows by rote that Geneva is in Switzerland, but he does not understand the spatial part-whole relationship. Probing questions reveal that he thinks he can be Genevan or Swiss but not both. Only between the ages of ten and eleven does a child grasp that Geneva is part of Switzerland.

## The Riddle of Space

Position in space is relative. Boston is north of New York. But New York, though south of Boston, is north of Washington, D.C. The child finds it difficult to understand these relationships in space.

Indeed, the child finds it difficult to understand space at all. Actually he does not believe in its objective existence but thinks that it disappears when he does not see it any longer. For a very young child riding on a train, scenery that disappears no longer exists, just as a watch hidden under a pillow before the baby's eyes ceases to exist for him.

Before children can understand spatial position, they must be free from personal, subjective perceptions and capable of co-ordinating spatial relationships.

### Direction

The seven- or eight-year-old child has learned to distinguish between right and left from his point of view, but he still believes that "right" and "left" are absolute properties. He does not realize that "right" for him is not the same direction as "right" for someone facing him.

After he has learned which of two objects is to the right and which to the left—that is, to his right and his left—he still does not recognize the relationship of the two objects to each other in respect to right and left. When he is about eleven or twelve years old, he can relate three objects to one another.

He understands that the middle one is left in respect to the one object and right in respect to the other object.

Seven- and eight-year-old children were shown a view of three mountains of different shapes and heights made of cardboard and standing next to one another. Then the children were shown several drawings of these mountains sketched from various perspectives. The children were asked to place a paper doll at points where the dolls could view the mountains from each perspective shown in the drawings. The children were unable to put the dolls in all the appropriate positions. They were able to place a doll so that it could look out on their own view of the cardboard mountains. Older children realized that the dolls' perspectives differed from their own but still could not analyze the relationships of the three mountains to one another.

At ten and a half years the child co-ordinates all the spatial relationships involved and achieves a true "decentration" because he pays attention to all possibilities, looking at the pictures from the point of view of the dolls and from the point of view of the three mountains.

Understanding of space develops gradually, as we have seen. The child must conceive of space as an orderly arrangement, consisting of an orderly distribution or expanse which is invariable, that is, permanent.

## Distance, Time, and Speed

Children younger than eight or nine cannot grasp the idea of velocity, because it involves both distance and time. Usually the child must be at least twelve before he can solve the more complex problems in this area.

Perception of time and distance is misleading even to adults, for personal experience does not create the feeling that time and distance are permanent. Time flies, and space passes. For the child not yet eight years old space is variable, impermanent; there is no orderly arrangement within space, no systematic relation of distances. When the child describes the buildings he passes on the way to school, his favorite buildings—like the candy store—seem to be much closer to home than they really are. All of us know how personal the flow of time is. An hour spent listening to a boring lecture seems longer than an hour spent looking at an interesting show. With the child an hour of rest time seems longer than an hour of play time.

Time must be regarded as an evenly flowing principle of chronological order, common to all, in which a succession of events occurs.

Velocity involves time, space, and movement. Time must flow evenly regardless of distances covered at different speeds. To understand movement, spatial order must be recognized, for movement is displacement in space covering a succession of intervals or distances.

Piaget's subjects had great difficulty relating time and space when they were solving problems on movement. The children focused on time or on distance but were unable to consider time and distance simultaneously. The idea of "before" and "after" in time created confusion with the "before" and "after" in space. In the experiment child and investigator walked a certain distance and dolls were moved through an unequal distance at the same speed or through the same distance at unequal speed. Children younger than seven or eight years of age thought that two people who arrived at the same point together must have walked at the same speed even though they had moved an unequal distance. Or the children believed that the person who had covered a shorter distance at the same speed and arrived sooner must have walked more slowly. The concept of velocity cannot be understood without "mental operation," which integrates two systems of reference—time and space.

## The Curriculum Reconsidered

Many of the concepts we have discussed are necessary in the sciences—such concepts as permanence of quantity, of size, of weight, of volume. These concepts, as we have seen, develop gradually from about six to twelve years of age. The concepts begin to develop at different ages, and they are grasped at different ages.

In the beginning school years, the young child perceives the world egocentrically and does not understand physical causality. At six or seven years of age the child has no true concept of time and space, of geographical relations, of past and future. Yet in many countries the curriculum and teaching methods fail to take this fact into consideration. School programs and teaching practice are not related closely enough to intellectual growth.

Five- and six-year-olds are often provided with school experiences designed to give them an understanding of distances and the relative speeds of walking and riding. Children of this age are given other experiences designed to teach them the meaning of weight and measurement, the meaning of relationships between the whole and its parts, the meaning of time. Historical events are discussed before the child understands present time, before he is ready to understand the past. Foreign cultures and remote geographical regions are studied before "far away" can be meaningful to the child, before he can understand the invariability—that is, the permanent organization—of space. Maps are shown to children at a stage in their development when a flat map cannot create a true picture for them.

## Mathematics . . . A Mystery Still

The greatest change that is needed seems to be in the teaching of mathematics. Because many children have found this subject difficult, schools in many countries have changed their teaching techniques. In many nations the study of mathematics used to start with number symbols. Now in early mathematics lessons the children count—first things and later pictures of things. Place-value charts are used to create an understanding of digits. Flannel boards are used to clarify the meaning of fractions and their comparative value. Some schools no longer teach mathematics directly in the early grades but use calculations in connection with the sciences whenever the need arises for counting and measuring.

True, children are more interested in calculating for a purpose than in calculating for the sake of learning. Yet nothing seems to reduce the number of failures in mathematics. All our tricks cannot provide understanding of general principles before the child's intellectual development has reached a point that will permit him to grasp them.

From the research of Piaget and Inhelder we know that a young child does not understand the cardinal value of a number, that is, the different possible combinations that will result in a given amount, nor does he understand the ordinal position of a number in a series. To the very young child, the amount seems to depend on the order of the objects. When he counts objects from right to left and then counts the same objects from left to right, he is surprised that the result of the two counts is the same. At the age of six he does not understand why this must be so.

## Lessons That Come Too Early

The processes of adding and subtracting cannot be clear until cardinal and ordinal concepts are understood, and this understanding develops at about seven and a half years of age. Before then digits and class inclusions are also beyond the child's grasp. As we have seen, the child does not reverse his manipulations to test his hypothesis; he does not separate a sum into its elements; he cannot establish relations beyond the one-to-one correspondence. When the appearance of a collection of objects changes, he thinks that the quantity also changes. According to Inhelder's research, one out of four children six and a half years of age still believed that if you moved objects closer together or farther apart you changed the number of objects. These children believed that beads increased in number when they were spread apart. Take a number of eggs and an equal number of egg cups. Push the cups farther apart or move them closer together and the number of cups is no longer the same as the number of eggs. So the children thought.

Despite what Piaget's research shows about children's thinking, in many countries addition and subtraction up to one hundred are taught in second grade, and multiplication and short and long division are taught in the third year of school.

Thus we often force the child to learn by rote or by manipulation lessons that come too early to impart an understanding of principles, too early to give the child insight.

Experiments and manipulation are excellent when the child is ready to understand them. Inhelder has found that children who are introduced to manipulation at an early age form basic arithmetico-geometrical ideas more easily. If, however, experimentation is introduced before the child can generalize, he will not gain insight. Given an appropriate experiment at the suitable age and not earlier, children can come to understand the concepts of invariability more easily, move through the stages of mathematical development faster, and deduce general principles more readily. In a study of the mathematical development of children from nine to eleven years of age, Inhelder discovered how much learning occurred while her subjects experimented with the materials presented to them. Manipulations greatly aided their insights at that age.

Basic geometrical concepts depend on an understanding of spatial relationships. Only when the child has liberated himself from subjective perceptions can he see the relativity of spatial positions. However, perceptions of spatial configurations are only the first step toward understanding of geometrical concepts. The concrete, specific character of spatial configurations impedes true mental operation and easily makes thinking about geometrical problems rigid. Perception alone—without liberation from the particular constellation—is irreversible, immobile, and static. Attention given to a certain spatial configuration limits thinking to the specific "picture" in view. A picture, a concrete perception, cannot be reversed; hence it encourages rigidity of thinking.

To think abstractly and to form generalizations, a person must lose his dependence on a concrete perception or representation. This liberation from the specific allows for mobility within the configuration. It makes mental displacement of the elements possible.

This mode of thinking—that is, conceptual reasoning without concrete materials—is typical of the adolescent. His thought is no longer a mere reproduction or an internalized image of reality but has become a reversible mental construct.

We educators face the problem of recognizing the kind of thinking the child is capable of at certain ages. Studies like those of Piaget help us realize the complexity of intellectual development. We have often assumed that a child could solve certain problems because he was able to solve others that appeared equally difficult to us. Searching for the causes behind the child's reasoning, Piaget shows why some problems require a more advanced mechanism of thought than others. Piaget's conclusions on the development of logical thinking and its specific forms at different levels ought to be of help to teachers in choosing methods of presentation. Piaget's findings may help curriculum-builders envision desirable changes, particularly in teaching the physical sciences, mathematics, geography, and history—subjects in which postponement of some areas seems desirable.

NOTE: For a searching and thorough analysis of Piaget's theory and research the reader is referred to John H. Flavell, *The Developmental Psychology of Jean Piaget*, Van Nostrand, Princeton, N.J., 1963.

# 8

# Education and the Psychology of Individual Differences — *Philip E. Vernon*

*There is a need for educational stimulation that is suited to children at all ability levels. Dependence on chronological age as the sole criterion for placing a child in a particular group at school often results in inefficient and frustrating learning situations, both for teachers and for students. In this article, the author discusses the many practices and proposals, current in both England and the United States, that are designed to provide for differences in intellectual ability among students.*

ONE OF the most urgent and most controversial questions in education today is what kind of organization will encourage the fullest development of the varied mental capacities and inclinations of students. There are certain fundamental psychological principles, and some advances in recent years in the field of mental testing, which may help to guide our views on educational policies. Although the present system of allocating students to suitable courses of study in England is about as different as it could be from the system in America, and although I possess little direct knowledge of the American educational system, it would seem worth while to pool our experiences and the results of the tremendous amount of research that has been done, in an attempt, perhaps, to see our way a little more clearly.

One of the chief advocates of educational acceleration and scholarships

Reprinted from *Harvard Educational Review*, 1958, 28, 91–104, by permission of the publisher.

for the gifted has been Professor S. L. Pressey (9) of Ohio State University. Let us take as our starting point three of the main arguments that he puts forward. First, the education of the most talented is unnecessarily lengthy, and, since adults tend to reach the peak of achievement in most fields in their thirties or earlier, we are wasting much of their productive life. Second, acceleration does not involve hurrying or strain, but represents rather the natural pace for the more able. Wherever it has been tried it has worked successfully, and there is no need for it to lead to any social or emotional maladjustment. Nevertheless objections of this kind are continually being put forward, when the real reasons for doing nothing are the economic and financial consequences, or pure inertia or prejudice. Third, a vast range of differences in abilities and interests exists among students who have had similar educational opportunities.

Now the problem is even wider than this. In general, pupils and students are so heterogeneous in their characteristics and potentialities that it might seem desirable to plan for each one an education uniquely suited to himself. However this is neither possible, nor desirable, for education implies not only individual development but also the training of different individuals to conform to society's patterns of intellectual and social norms. Thus there are positive advantages in educating diverse individuals in groups. Nevertheless, there must be some restriction of their heterogeneity, otherwise the educational process becomes inefficient and frustrating to the students as well as to the teacher. It would be absurd, for example, to try to train imbecile children and university students of atomic physics in the same group. To take a less extreme example: the tiny country school where one teacher copes with a very wide age and ability range performs many valuable functions; but there is no doubt, in England at least, that its educational efficiency tends to be below average.

Given, then, that there must be some reduction in heterogeneity, the psychologist would surely urge the following stipulations. Any grouping should be based on some characteristic which: first, is stable and enduring; second, can be accurately assessed; third, has a major influence on educational progress; and fourth, is acceptable to society. Difficulties arise because few characteristics, apart from age and physical handicap, meet these requirements. Age does largely, though of course far from completely, determine intellectual, emotional, and sensory-motor maturity; there are no difficulties of assessment, and it wins general acceptance. There is general agreement, also, that the deaf, the blind and partially sighted, and certain other physically handicapped groups should be segregated for special schooling under specially qualified teachers, though doubts arise in deciding what degree of defect requires such treatment. All other types of homogeneity seem to arouse intense controversy. Differentiation on the basis of sex, for example, is rejected by most psychological opinion; yet it occurs in many European and a few American schools and colleges and is strongly supported by many parents and alumni. The explosive topic of race will not be considered, and religion will be omitted as having relatively little relation to educational capabilities. Socio-economic class deserves some consideration.

Social class is fairly stable and assessable. We know that it is so closely associated with cultural level and with attitudes toward education that it has a marked effect on educational progress. In England, for example, where children are graded according to their suitability for advanced secondary education at eleven to twelve years of age by means of objective tests of intelligence and achievement, about three times as large a proportion of children of the white-collar classes pass the tests as do children of manual workers. Furthermore, social class gives one of the best indications of future achievement and adjustment in school. Havighurst and his colleagues have shown that, despite the greater social mobility in America, the social class of an American student's parents largely determines the stage he reaches in the educational ladder, and the kinds of courses he selects in high school (*16*). Thus there is some justification for the practice, common in Europe and not unknown in the United States, of having separate secondary schools for the middle and upper classes, entry to which is mainly by parental income. This system, however, is repugnant to the temper of the age in all democratic countries. Unfortunately, class is so pervasive that almost any form of selection or grouping is likely to be affected by it. Thus, selection by ability has become a source of acute political dissension in England between the more conservative who wish to retain it and the socialist or working class who wish to abolish it and substitute something more like the American common, or comprehensive, school.

Yet another, apparently sensible, criterion for the grouping of pupils—namely, a common linguistic background, usually breaks down because it too is associated with racial or with social class differences. One important reason why upper class parents tend to segregate their children, though they would seldom openly admit it, is that they do not want them to talk like children in the industrial slums. The psychologist may deplore these sociological and political influences on education, but he cannot afford to neglect them.

Let us turn then to ability grouping. In the early 1920's, when group intelligence tests became widespread, it was all very simple. We were taught that the tests measured innate intelligence, unaffected by social background, and that this was the main determiner of educability. Thus if pupils could be put into groups which were homogeneous with respect to intelligence, each group could progress at its own rate and the teacher's burden would be very much lightened. Moreover, the IQ was supposed to give a reliable indication of a child's ultimate educational powers; whether he was bright enough, or too dull, ever to manage work of college level, or to tackle advanced topics like mathematics. To a large extent this advice given by psychologists was put into practice in England. Age grouping gave way to ability grouping, though it was soon realised that too wide an age range of bright youngsters and old dullards in a single class is socially unhealthy. The usual practice, then, came to be the classification of children within an age grade in any large school into three or more streams or tracks on the basis of intelligence, previous achievement, or a combination of the two. Also, as mentioned earlier, the brighter children were segregated from the average and duller ones at eleven years for accelerated schooling.

Nowadays views on intelligence have been greatly modified. Indeed the term has been largely discarded by many American psychologists, as too liable to misinterpretation. Nevertheless, the distinction between intelligence and attainments still has limited value, despite their close overlapping. Intelligence can be defined as the more general thinking capacities: capacities for reasoning, for grasping relations, for comprehension, for new learning, and for concept development. That is, capacities which are not so much specifically taught as picked up by children in the course of their interaction with the home, school, and wider environments. Whereas attainments refer rather to concepts and skills which depend more on direct instruction and on the child's interest and industriousness in the particular subjects studied. Such a distinction is one of degree rather than of kind, and it is entirely false to think of intelligence as causing, or making possible, attainments. It is at least as arguable that, through the acquisition of attainments at home and at school, the child is enabled to build up his intelligence.

Both depend to a considerable extent on innate potentialities or maturation—on what the Canadian psychologist Hebb (4) calls Intelligence A—some characteristic of the central nervous system which enables certain children to develop mentally, to form percepts and concepts, habits and ideas, to build up complex intellectual skills more readily than other children. Intelligence A is the capacity to acquire Intelligence B—that is, the intelligence which we actually observe in everyday life or at school, and which our tests sample fairly effectively. Intelligence B, as Jean Piaget (8) also shows in his post-war books, is built up gradually; it does not depend solely on the child's genes, but also on the stimulation of the child by the world in which he is reared and which gets him to exercise his potentialities. How far specific attainments involve different genes from those underlying general intelligence seems to me a moot point; one could certainly make a good case for it in the field of music and possibly other talents. But, like most heredity-environment controversies, this seems an unprofitable argument, since we can never in fact observe or measure Intelligence A or other potentialities directly; they are purely hypothetical constructs.

We can no longer regard intelligence as setting a fixed rate of mental growth, or as having a definite termination. The early pioneers of mental testing seem to have regarded the IQ as a fairly constant characteristic of each individual because there was good agreement or correlation between two applications of the Stanford-Binet test a week or a month apart. But this tells us little about the fluctuations in ability to be expected during five or more years of schooling. On surveying the many investigations that have been made into this problem recently, a number of technical snags were found. Most of the published results were more or less distorted by such factors as abnormal standard deviation of the IQs, extreme similarity or dissimilarity of the tests employed, above average intelligence level of the tested group, frequency of retesting, etc. However, allowing for these it appeared safe to conclude that either over the six to ten or the eleven to eighteen year period, the correlation between two similar (though not identical) intelligence tests does not drop below a coefficient of 0.70. This implies that the typical individual would vary only 7 IQ points up or down.[1] Although the majority do not fluctuate very widely, as many as 17 per cent might alter considerably, that is, anywhere from 15 points (or 1 standard deviation) up to about 40 points.

All-round educational attainments seem to be at least stable. Certainly it is a mistake to think of the IQ as something fixed, and attainments as varying widely with the teaching received and other environmental causes. I would suggest that both show moderate constancy, partly because both are based on genetic potentiality—Hebb's Intelligence A—but also because their growth is essentially cumulative. By the time the child reaches school at five or six years, the interaction of potentiality and home stimulus have consolidated in him a certain level of ability which will closely determine his rate of progress for the next few years. Good or poor teaching, or other marked environmental irregularities and changes in personality adjustment, motivation, and interests, will result in considerable alterations only among a small minority. Again by eleven or twelve years the consolidated level gives fairly close predictions of the rate of progress over the next six years, as we have been able to show by our follow-up studies of the English secondary school examinations (*15*). The correlations of intelligence tests, of English and Arithmetic tests, and of teachers' estimates of ability with successful secondary school performance over two, and even up to five years average close to 0.80, when suitably corrected for homogeneity. By combining these three sources of information the coefficient can be raised to about 0.86. This, of course, refers to a whole age group. The figure may sound surprising to American educational psychologists, because they normally do their testing within pre-selected groups such as students seeking college admission. But if they likewise calculated the efficiency of their aptitude and achievement tests and school grades in separating, say, the top 20 per cent of high school students who are most suitable for college courses from the bottom 80 per cent, there is no doubt that they would reach at least as good predictions.

---

[1] The statement assumes that such tests are standardized to the same mean of 100 and standard deviation of 15.

Yet even a correlation of, say, 0.85 allows of a considerable degree of error. Suppose we did send 100 unselected eighteen- to nineteen-year-olds to college and found which twenty were the most successful, in fact only fourteen of the twenty would have been correctly indicated by our tests and high school grades. In other words, using the best criteria available, about one-quarter to one-third of those we should select for college are likely to be unsuitable. They would have been equalled or surpassed by some 7½ per cent of those thought unsuitable, had the latter been given the opportunity. We can represent the situation as follows:

| | Successful in College | Unsuccessful in College | Total |
|---|---|---|---|
| Selected on the basis of tests and previous grades | 14 | 6 | 20 |
| Rejected on the basis of tests and previous grades | 6 | 74 | 80 |
| Total | 20 | 80 | 100 |

This is precisely the situation that obtains in English selection for advanced schooling in what we call grammar schools. A very high statistical level of accuracy is reached, and yet many pupils admitted to grammar school are unsuitable, and some who are rejected turn out to be so-called late developers who have to be transferred later on, or otherwise show outstanding ability. The psychologist, however, would naturally expect such fluctuations in abilities during adolescence and young adulthood; as students progress to more advanced courses, their interests and adjustment often change and they develop unsuspected capacities, or drop below their previous promise. Whatever kind of system of selection or acceleration was adopted in America, really accurate prediction would likewise be found impossible.

Frequently one hears the argument that far too large a proportion of our talented youth fails to get to college; there are many millions with high IQs who are being wasted. This argument is no more impressive than the statement: since college students are above average in height, we should ensure that all six-footers reach college. While IQ may be somewhat more relevant, its correlation with success in college for a representative sample would not exceed 0.75, and this means that 40 per cent of students picked merely by high IQ would turn out unsuitable. Even if we raised our standards and insisted, say, that all students with IQs of 125 and above, the best 5 per cent only, should be accelerated in high school, or sent to college, nearly one-quarter of our choices would let us down. At the opposite extreme, even if we went as low as an IQ of 90 and included the top three-quarters of the population, we would still find that 2 per cent of those excluded might have made the grade. In order to catch all or nearly all those capable of benefiting from a more advanced or accelerated educational course, we should need to go so far down in the scale of ability that many students would be brought in who would not benefit or who might have been more appropriately allocated to some quite different course, say of vocational training.

No definite conclusions or recommendations will be arrived at in this paper; but the main educational implication so far would seem to be that any predictions we make about students, or any selection, should be short-term and flexible. Such procedures should take into account not merely abilities but also interests and values—whether, for example, the student and his family are keen for him to take college preparatory courses or to go to college. While all assessments are likely to have a considerable margin of error, they will certainly be less inaccurate when based on the way the student is shaping and progressing in work of a similar nature to that which he will undertake. Clearly, too, caution should be exercised regarding any statements which imply that a considerable proportion of the population is always going to be intellectually incapable of absorbing the higher reaches of education. There is still a lot of truth in this, but the trouble is in determining *which* members of the population fall into this category. The limitation is at least as much a matter of defective intellectual interests and defective attitudes to education in the environment from which the weaker students come as it is of these students' innate intellectual inferiority.

Another deduction may be drawn from the work of Hebb, Piaget, and others on intellectual growth: such work suggests the desirability of the educational environment keeping just sufficiently ahead of each pupil's capacity to stretch his mind to the utmost. Hebb found that dogs or rats brought up in the restricted environment of a cage were less able, as adults, to perform new learning and problem solving tasks than those brought up in a richer and freer environment. Similarly with children, inadequate educational stimulation may mean not only that they learn less than they should, but that they also become less able to acquire further thinking skills. Experimental evidence can be provided on this point.

Recently, some 800 boys in a large English city who had been tested and allocated to fourteen very diverse secondary schools at age eleven were retested at fourteen (*14*). After allowing for initial level and regression effects, there were differences between the different school groups amounting to 12 IQ points. The pupils in the average selective or grammar school had gained 7 points (nearly half a standard deviation) over those in the non-selective schools. All this difference cannot be attributed to the effectiveness of schooling as such. Grammar school boys mostly came from better homes where they received more encouragement and help with their work, while the non-selective school boys lived in culturally and materially poorer homes, where there was often active opposition to education. Nevertheless, a rank correlation of 0.85 was found between the city administrator's assessment of the degree of stimulation likely to be provided by the fourteen schools, and the actual order of average gains among the pupils in these schools. Clearly, much more transfer had occurred than Thorndike found in his classic experiment on different high school courses (*10*).

This finding recalls Lorge's work on the intelligence level of thirty-four-year-old adults who had varying amounts of secondary and tertiary education after they were tested at fourteen years (*6*); and also a recent study by Husen of several thousand Swedish boys at nine and nineteen years. Husen found that those who had had full education throughout the period gained the equivalent of some 12 IQ points over those who had left school at fourteen to fifteen and received no further education (*5*).

During World War II, under the auspices of the British Navy, the Raven Matrices scores (a non-verbal reasoning test) of 90,000 men recruited at various ages from seventeen to forty were compared and a general tendency to decline with age was noted (*13*). Clearly, however, those who had been in unskilled and laboring occupations, where they had made little use of their brains, declined earlier and more rapidly than those from skilled trades and clerical work, which had presumably done more to exercise them.

Now according to the older views, intelligence was said to grow steadily in childhood, then to slow down and reach a constant maximum at around fourteen to fifteen years of age. No one ever explained how, if this were so, high school and college students could continue to increase their scores on suitable tests in the late 'teens. But these results are quite reconcilable if we remember that, around the time of World War I, probably the majority of average and duller individuals were leaving school by the age of fifteen; their intellectual capacities not only ceased to develop but began to stagnate, whereas the brighter, who either stayed on at school or entered more stimulating jobs, continued to improve. Thus the combination of a rising group and a declining group produced the apparently constant average level in the adolescent and young adult population.

The objection may be raised, as it was raised to Lorge's findings, that intelligence tests depend greatly on vocabulary and reading comprehension, capacities which are naturally much affected by length and quality of schooling. The objection rests, of course, on the old confusion between intelligence as a hypothetical innate potentiality (Intelligence A) and intelligence as directly manifested (Intelligence B). Actually, further work by Lovell, indicates that it is not mere verbal skills which are most affected by the stimulating or depressing qualities of adolescent environment, but rather the flexibility aspect of intelligence, or the capacity to form and apply new concepts (*7*). A more serious weakness is that we know so little, as yet, regarding the kinds of educational process that have greatest transfer value, or do most to stimulate the growth of ability.

Now the implications of such findings for educational organisation would seem clear. They definitely support some system like the English one of grouping pupils by ability, and pushing forward the brighter ones more rapidly in selective schools or classes. And they suggest that there is considerable substance in the complaint made by many American parents that the American system of public education seriously retards the bright child who is willing and able to tackle more difficult courses. On the other hand, there is the danger that average or duller students, if they are less pushed, will tend to fall more and more behind. Thus, it becomes even more difficult for those whose capacity happens to improve later to get back into the stream that would be appropriate. In other words, homogeneous grouping by ability tends to stereotype those who were less able initially and freeze them at a lower level.

This is a very real problem in English elementary schools, where pupils may be classified by teachers as early as seven years old into those thought likely to pass or fail the eleven year examinations. Such classification, being based mainly on early progress in reading and number-work, naturally gives a great advantage to children from middle class families who are more likely to be helped and encouraged at home. Thus, a rather rigid system of selection may mean that the initially slow learners drop so far behind that they are, in effect, being condemned very early to an inferior schooling and a manual occupation for life. However, evidence indicates that such rigidity is the exception rather than the rule; perhaps owing to the influence of psychological teaching, most primary schools now do retain a fair degree of flexibility of transfer from one stream to another, and a sufficient degree of overlapping between the curricula of the several streams (*1*).

A similar danger arises at the bottom end of the ability range, where, in many educational systems, the very backward pupils may be classified as mentally defective and sent to special schools. It may be that their abilities too become stereotyped at this level, although they might have been capable of returning to ordinary school classes later. Miss Bernardine Schmidt's famous research at Chicago definitely supports this view, although her claims are by now pretty thoroughly discredited. In England, at least, most mental defectives receive a type of education far better suited to their lowly intellectual level and their weak emotional adjustment than they would get in ordinary schools. Yet they usually stay as backward as ever or even decline. No one who has observed such children vegetating hopelessly in an ordinary school, and then busy and interested at a special school could doubt the desirability of some type of homogeneous grouping. Nevertheless, I agree that here too irreversible segregation is to be avoided, and that injustices do occur.

One other consideration that I wish to raise in respect to ability grouping is that—as educationists well know—any scheme for grading of students has a "backwash" effect on education, by acting as an incentive to the teachers, the students, and their parents. The eleven year selection examinations in England provide a terrible object-lesson; many, though by no means all, primary schools concentrate so exclusively on cramming their fifth and sixth grade pupils for objective tests that any other educational activity, however valuable it might be to children's general growth, tends to get crowded out. Much ill-feeling is engendered among snobbish parents; often they coach children themselves from published manuals of intelligence and other tests; or they communicate their anxieties to their offspring and sometimes induce serious strain. Actually the effects on young children's mental health are apt to be greatly exaggerated by the sensational press. A careful survey of child guidance clinic cases has been made and no tendency for an increase in referrals was found around the time of the examination. In only one child in about 500 does the examination appear to contribute to maladjustment, and then only in children who were already prone to anxiety through earlier upbringing or constitutional weakness. Probably other competitive examinations, such as those at the end of a secondary course, are associated with mental breakdown at least as frequently. Nevertheless, the existence of unnecessary stress even in a small minority, and distortion of the educational process among the majority, are the most serious defects of the English system, and have to be weighed very carefully against possible advantages in intellectual acceleration.

Now although sound psychological reasons for expecting intellectual benefits through grouping can be given, it is very difficult to prove. Furthermore, such limited experimental comparisons as have been carried out seem to have yielded negative results. It is doubtful whether such studies have been adequately designed to answer the important questions. International comparisons, again, yield no sure evidence. It is commonly stated that the products of English grammar schools, when they enter college, are two years ahead of American students of the same age. Even if this generalization were true it proves nothing, since only some 3 to 4 per cent ever reach the university in England. Also, it may well be that superiority in academic achievement is counterbalanced by poorer development in social and other more subtle qualities. It would seem, and some day it may be shown empirically, that the English system of grouping and selection does produce improvement in achievement at the top end, whereas the common school system is better for the average or dull, and also obviates many of the difficulties that arise through fluctuations in, or stereotyping of, mental growth. If this were so, it would be a matter for society rather than psychologists to decide which outcome it prefers.

No doubt it will be suggested that, instead of grouping by general intelligence and previous all-round achievement, we should consider type of ability and interests along special lines. This might be called the multi-dimensional as opposed to the uni-dimensional approach. British and American psychologists have always differed on this, just as have their educational systems. The British, following Spearman, have tended to stress the importance of general ability: children who are above average in one school subject are likely to be above average in all others. Only to a limited extent, as Burt demonstrated in 1917, do we get more specialised types of ability showing in the primary school—for example, some children being generally poorer at number work than at English subjects, or better at all practical and manual activities (*2*). In contrast, Thurstone in 1938, from his analysis of tests given to college students, claimed that abilities are largely independent—attributable to a series of separate factors; verbal, number, reasoning, spatial, etc. When he extended his testing downwards to eighth grade and younger children, who were less highly selected, he in fact found much overlapping among these factors (*11, 12*). So that, while his statistical approach was quite different, his results confirmed Burt's in showing that something like 50 per cent of the variance in children's abilities, at least in abilities that are relevant to educational progress, could be accounted for by a general factor. This means then that it is extraordinarily difficult to differentiate secondary school pupils by type of ability into even such broad categories as linguistic or academic on the one hand, and scientific or technical-mechanical on the other. For the most part, those who are best at one course would also be above average at the other, and vice versa.

If we confine our attention to a narrow ability range, say the top 20 per cent as in England, then we find it possible, with our present verbal and spatial tests, to differentiate the extremes. Of course there are no distinct types of children any more than there are tall and short types in height. We can pick out roughly one-sixth who are relatively superior on the linguistic to the technical side, and one-sixth who are more suited to the other type of course. There remain two-thirds who are likely to do equally well or badly at either. Roughly the same is true of American college students. The efficiency of prediction of overall level of achievement by means of tests is much greater than the efficiency for classifying according to, say, engineering or arts, and this is only the broadest and most obvious dichotomy. If we tried to go further and split off, say, foreign linguists, social studies students, mathematicians, biologists, physicists, engineers, commercial students, art students, and agriculturists, we should certainly be far less successful still. Probably the linguists, mathematicians, and physicists would be distinguishable from the artists and agriculturalists more readily by their generally higher intelligence and previous attainment than by such specialised tests as we have available at present.

Some psychologists would claim that abilities differentiate with age, so that suitability for different lines of study or type of curriculum would become more clear-cut near the end of secondary schooling or during college years. No satisfactory evidence for this claim can be found except in so far as groups of older students are usually more highly selected in respect to general ability. In the independent college, the scientist might make a hash of French, or the good historian fail in engineering, because in such a population practically all cases fall within the top 10 per cent as regards general ability. There would be much less differentiation in the state universities, where students range roughly over the top 30 to 40 per cent of the scale.

However, the outlook for guidance or selection into types of courses is not quite so dim if we take account of interests, since interests have the tremendous advantage, for classification purposes, of showing quite low, sometimes even negative, correlations with one another. The good scientist and the good business man, for example, may differ little in such abilities as we are able to test. They would however loathe one another's jobs, and would likely be thoroughly bored by any education or training designed to prepare them for the other's technical career. In England we find that we can differentiate nearly two-thirds of the more able pupils into the academically- and the technically-minded at eleven as against the one-third mentioned earlier, by taking account of *interests* as well as *abilities*. One difficulty with interests, though, is that they cannot develop without experience. There must be many adults who fail to get much satisfaction from their vocational and avocational pursuits because their education has not provided sufficient opportunities for them to experience, say, painting, farming, or mechanical work. The American system of offering a wide choice of secondary school subjects, while often criticised by educators for its fragmentariness, is sounder in this respect than the English which gives less variety and indeed tends to stereotype the more able pupils either in an Arts or a Science field from about the age of fourteen, and to discourage less academic inclinations. The shortage of technologists and high-grade technicians in England is at least as serious as that in America. Clearly one of the reasons is the high social prestige of academic courses of study which hinders the provision of experience that might stimulate the development of more practical interests.

Another obvious difficulty arises from the variability of interests during adolescence. Among many students interests may not be sufficiently stable nor assessable to allow firm and satisfactory educational or vocational specialisation until the end of secondary schooling or even later. Nevertheless, one investigation at London University of adult technologists and Arts men did suggest that some differentiation of interests would have been apparent in the great majority as early as twelve years (3). This is a field where psychologically trained counselors can already help in making earlier diagnoses, and where current work on personality gives promise of improved tests.

How about the problem of discovering the brilliantly talented individual—the occasional future genius? Clearly, neither our tests of abilities nor of interests are going to be much use here, and we must continue to trust mainly to luck, to the perceptiveness of the school or college teacher who encourages him, and to the individual's own drive, which enables him to break through the barriers of the educational system. Admirable as Terman's follow-up studies of children with high IQs are, some harm has been done by labelling them studies of "genius."

One other point of psychological theory: although it has been shown that the group factors underlying particular types of curriculum are disappointingly small and difficult to diagnose, this does not dispose of what Spearman called specific factors. Individual pupils show great variations with respect to specific topics within any course, depending on their particular past experience and interests, and on their emotional responses to the teacher and other members of the class. Thus, even when they are satisfactorily grouped for, say, oral reading, they will show considerable heterogeneity in silent reading or spelling, let alone in non-related subjects. For such reasons, many educational psychologists currently look more to individual assignments and small-group work within classes to cope with most of the problems of individual differences than to overall grouping into separate classes. With this goes the notion of enrichment of courses among the brighter students. The adequacy of such solutions is doubtful since they clearly imply increasing the degree of heterogeneity within the class and thus run counter to the thesis of this paper.

What conclusions can we come to then? No evaluation of the various mechanisms of acceleration that have been, or are being, tried out in America such as grade-skipping, advanced standing, and the shortening of courses through enriched programmes will be made here. But the evidence does seem to point to the desirability of some form of grouping, such as indeed already exists in many American high schools. Nevertheless, we have seen that there are many dangers in introducing anything that implies competitive selection or stereotyping of ability levels. It is, therefore, preferable to keep to grouping by age, and later by interest, as far as possible. Up until about nine years of age or the fourth grade, there would seem to be no good case for any ability grouping other than segregation of the lower-grade feebleminded, the physically handicapped, and, perhaps, temporary remedial classes for the higher-grade defectives and the very backward. However, by the age of about thirteen or the eighth grade, the range and complexity of abilities *appears* to have become much wider. It is still doubtful how satisfactorily we can measure range of ability in any absolute sense even at this age. But everybody can't go on studying everything, and, in the interests of professional and vocational objectives, some specialisation should begin to be introduced. Hence, some degree of grouping by general ability, and to an increasing extent by interest, would seem legitimate. While it is true that many thinkers strongly deprecate early specialisation, surely its dangers should be balanced against the waste of productive years which results at present from the late postponement of any specialisation. Between the fourth and eighth grades one can only suggest that the curriculum be largely of an exploratory or diagnostic character, designed not merely to provide essential skills, but to stimulate general mental development and to provide experience out of which interests can be built. Even when grouping is introduced, there should be sufficient overlapping between groups to make transfer up or down easy, and of course sufficient common activities in a school or college to discourage the formation of barriers. In other words, the process should be one of gradual approximation in accordance with the principles of educational guidance rather than one of selection or irreversible decision.

Inevitably this is a vague kind of framework and the psychologist must always remember that his prescriptions are liable to be upset by social prejudices and traditions, financial shortages, increasing birthrates, and innumerable other factors. The American type of school organisation seems to meet the prescription set forth in this paper more nearly than does the British, though it may, of course, show weaknesses in other respects. Some means should be devised of giving greater and earlier recognition to individual differences in general educability and of avoiding, if possible, some of the mistakes that have arisen in England as a result of the strangle hold of tradition and the well-intentioned but short-sighted policies of educational planners.

## REFERENCES

1. Blandford, J. S. Standardized tests in the junior school with special reference to the effects of streaming on the constancy of results. *Brit. J. educ. Psychol.*, 1958, 28.
2. Burt, C. L. *The distribution and relations of educational abilities.* London: King, 1917.
3. Clark, J. F. A psychological and sociological study of students in higher technological education. Unpublished Ph.D. thesis, Univer. of London, 1951.
4. Hebb, D. O. *The organization of behavior.* New York: John Wiley, 1948.
5. Husen, T. The influence of schooling upon IQ. *Theoria,* 1951, **17**, 61–88.
6. Lorge, I. Schooling makes a difference. *Teach. Coll. Rec.*, 1945, **46**, 483–492.
7. Lovell, K. Intellectual deterioration in adolescents and young adults. *Brit. J. Psychol.*, 1955, **46**, 199–210.
8. Piaget, J. *The psychology of intelligence.* London: Routledge & Kegan Paul, 1950.
9. Pressey, S. L. Educational acceleration: appraisal and basic problems. *Ohio State Univer. Bureau Educ. Res. Monogr.*, 1949, No. 31.
10. Thorndike, E. L. Mental discipline in high school studies. *J. educ. Psychol.*, 1924, **15**, 1–22, 83–98.
11. Thurstone, L. L. *Primary mental abilities.* Chicago: Univer. of Chicago Press, Psychometr. Monogr., 1938, No. 1.
12. Thurstone, L. L. and Thurstone, T. G. *Factorial studies of intelligence.* Chicago: Univer. of Chicago Press, Psychometr. Monogr., 1941, No. 2.
13. Vernon, P. E. The variations of intelligence with occupation, age, and locality. *Brit. J. Psychol. (Statist. Sec.),* 1947, **1**, 52–63.
14. Vernon, P. E. Intelligence and intellectual stimulation during adolescence. *Indian J. Psychol.*, 1957, **2**, 1–6.
15. Vernon, P. E. (Ed.) *Secondary school selection.* London: Methuen, 1957.
16. Warner, W. L., Havighurst, R. J., & Loeb, M. E. *Who shall be educated?* New York: Harper, 1944.

# 9

# Effects of Accelerating Bright Older Pupils from Second to Fourth Grade — *Herbert J. Klausmeier and Richard E. Ripple*

*One method of providing for differences in intellectual abilities among students is through acceleration. Many questions surround the practice of acceleration, i.e., of allowing the more able students to skip grades in elementary school. This article reports an experiment conducted to ascertain the effects of accelerating from second to fourth grade, pupils of superior learning abilities. These pupils were above the median chronological age of second graders, and in order to prepare themselves, they had attended a five-week summer session. Note the extent to which the authors have gone in attempting to account for the many variables that must be considered here, e.g., the number and kind of control groups, assessments made, etc.*

THE PUBLIC elementary schools of Racine, Wisconsin, are organized into age-graded classes—kindergarten through Grade 6. With few exceptions the children are at least 6 years of age on or before December 1 when entering the first grade. Children who are not subsequently accelerated range in age from 17-7 to 18-7 when graduating from high school, and 21-7 to 22-7 when completing four years of college.

Reprinted from *Journal of Educational Psychology*, 1962, 53, 93–100, by permission of the senior author and the American Psychological Association.

Since there are as many bright children in the older half of all children entering the first grade as there are slower learning children in the younger half, the question is raised:

> Should not the older bright children be accelerated at some grade or school level so that they may complete high school at least as young as their less able classmates who happened to be born during a month which permitted them to be the younger children when entering the first grade? (Klausmeier, 1958, p. 41).

The purpose of this controlled experiment was to ascertain the effects of accelerating from the second to the fourth grade pupils of superior learning abilities who were above the median chronological age of all second graders in the Racine schools.

## Procedure

### Subjects

In March 1960, 32 girls and 20 boys in the upper half of their grade in chronological age and of superior learning abilities were identified from the entire second grade population of the Racine, Wisconsin, schools. These pupils were born after June 1, 1952, had a minimum Kuhlmann-Anderson IQ of 115, a minimum total standard score of 300 on the Elementary Battery of the Metropolitan Achievement Tests, and a recommendation for acceleration from their teachers. The 52 pupils were ordered in pairs, matched by sex, and then randomly assigned, one from each pair to the accelerated group, the other to the control group of nonaccelerates. The accelerates attended a 5-week summer session in 1960 and were enrolled in 16 different fourth grade classrooms in September 1960; pupils in the control group did not attend the summer session and were enrolled in 15 different third grade classrooms in September 1960. The accelerates attended the summer session in order to complete the essential content of the third grade which had not been included in the second grade. Cursive handwriting, language arts, arithmetic, and expressive abilities—verbal, artistic, musical, and physical—received main attention. The accelerates met with their teacher in a self-contained classroom from 8:00 to 12:00 A.M., Monday through Friday.

Five additional groups were identified, each with 16 girls and 10 boys, for comparison with the accelerates. These were as follows: one group of third graders of superior learning abilities below the median age of third graders; two groups of fourth graders of superior learning abilities, one above and the other below the median age of fourth graders; two groups of fourth graders of average abilities, one above and the other below the median age of fourth graders.

The younger third graders of superior learning abilities were born after

June 1, 1952; they met the same criteria of IQ and achievement as did the pupils randomly assigned to acceleration or nonacceleration. Pupils in the two fourth grade superior groups had a minimum IQ of 115 and a minimum total standard score of 400 on the achievement battery. Pupils in the older group were born before June 1, 1951; pupils in the younger group were born after June 1, 1951. Pupils in the two fourth grade average groups had an IQ between 90 and 114 and a total standard score between 300 and 400 on the achievement battery. Pupils in the older group were born before June 1, 1951; pupils in the younger group were born after June 1, 1951.

Pupils in the four fourth grade groups had taken the IQ and achievement tests as third graders in March 1960, the same month during which the other three groups did as second graders. Pupils in the seven groups were enrolled in 16 different third and fourth grade classrooms in September 1960.

Comparisons between the two groups, originally matched by sex and then randomly assigned to acceleration or nonacceleration, were made on all data obtained in March 1961 (at which time the accelerated pupils had been in the fourth grade for 7 months of the 10-month school year). These comparisons were expected to demonstrate the effects of acceleration. Comparisons between the accelerated pupils and pupils in the other five groups were made to ascertain the extent to which the accelerates differed from younger third graders of superior learning abilities, and from older and younger fourth graders of average and superior learning abilities.

## Types of Data Gathered and Measuring Instruments

Nine types of data—educational achievements, attitudes toward school and learning, problem solving ability, ethical values, handwriting skills, psychomotor abilities, intellectual and affective characteristics, peer acceptance, and creative thinking abilities—were secured on pupils in March 1961. A brief description of the instruments used to obtain these data follows.

*Educational Achievements.* The Metropolitan Achievement Tests, Intermediate Battery-Partial were administered to all pupils. Scores were obtained for Word Knowledge, Reading, Spelling, Language Total, Language Study Skills, Arithmetic Computation, Arithmetic Problem Solving and Concepts, and Total Battery.

*Attitudes toward School and Learning.* A ranking scale and an inventory were devised by the writers to assess the pupils' attitudes toward school and learning. The ranking scale, titled Places Liked Best, required the pupil to rank the school classroom and eight other places from highest to lowest on the basis of preference for the places. The Attitude Inventory consisted of 20 multiple-choice items.

*Problem Solving Ability.* A test of 25 multiple-choice items was devised by the writers to assess a composite problem solving ability as represented in items dealing with analogies, logical reasoning, pertinent questions, sensitivity to problems, and following directions.

*Ethical Values.* A multiple-choice inventory was devised by the writers and research assistant, William Franzen, to assess ethical values. Responses were keyed as to whether they indicated rational conscientiousness, irrational conscientiousness, group conformity, or self-centeredness. The inventory yielded a score for each of these four traits.

*Handwriting Skills.* A speed score was obtained by having each pupil write a standard sentence as rapidly as he could for 3 minutes. The score was a count of the number of words written. A legibility score was obtained by having each pupil write the standard sentence once at his normal speed of handwriting. This sample was rated by three judges on a nine-point scale. The procedures of Virgil E. Herrick, University of Wisconsin, for administering and scoring these tests were followed.

*Psychomotor Abilities.* For tests of psychomotor abilities were devised and administered by Grace Piskula, Physical Education Supervisor of the Racine public schools: Basketball Throw to measure strength and coordination of body and arms, Jump and Reach to measure coordination of arms and legs, Wall Pass to measure eye-hand coordination and speed of reaction, and Shuttle Run to measure running speed and maneuverability.

*Intellectual and Affective Characteristics.* A Teacher Rating Scale was devised by the writers consisting of 12 descriptive items. The teacher rated the pupil on each item, using a four-point scale. Weights of one to four were given to each item and a total score was obtained by adding the weighted scores. In addition, three items dealing with the child's emotional, social, and physical development were checked by the teacher to indicate normal or unusual development.

*Peer Acceptance.* A sociometric instrument was used in which the pupils listed their five best friends in rank order. A score for each pupil was obtained by tabulating the choices received, weighting the choices, five for a first choice, four for a second choice, etc., and adding the weighted scores.

*Creative Thinking Abilities.* Seven tests of creative thinking yielding 10 scores were administered to all pupils: Object Uses, Word Uses, Plot Titles, Expressional Fluency, Plot Questions, Object Improvement, and Sentence Improvement. These tests, except Sentence Improvement, were adapted from Guilford (Guilford, Kettner, & Christensen, 1956) by research assistant, Frank B. May, for use with children. Sentence Improvement was devised by May. The reliability and validity coefficients for these instruments are reported by Ripple (1961) and May (1961); reliability coefficients only are reported in Table 1.

TABLE 1. Reliability Coefficients of Group Tests

| Test | Basis or Source of $r$ | $r$ |
|---|---|---|
| Metropolitan Achievement Test-Intermediate Battery: | | |
| Word Knowledge | Test manual | .94 |
| Reading | Test manual | .90 |
| Spelling | Test manual | .92 |
| Language Total | Test manual | .89 |
| Language Study Skills | Test manual | .79 |
| Arithmetic Computation | Test manual | .88 |
| Arithmetic Problem Solving and Concepts | Test manual | .92 |
| Social Studies Study Skills | Test manual | .73 |
| Places Liked Best | Alternate forms of test | .93 |
| Attitude Inventory | Cronbach Alpha | .79 |
| Creative Thinking Battery: | | |
| Object Uses-Fluency | 2 independent subtests | .77 |
| Object Uses-Flexibility | 2 independent subtests | .41 |
| Word Uses-Fluency | 2 independent subtests | .71 |
| Word Uses-Flexibility | 2 independent subtests | .66 |
| Plot Titles-Fluency | 2 independent subtests | .92 |
| Plot Titles-Cleverness | 2 independent subtests | .60 |
| Expressional Fluency | Cronbach Alpha | .76 |
| Plot Questions | Cronbach Alpha | .79 |
| Object Improvement | Cronbach Alpha | .29 |
| Sentence Improvement-Metaphors | Cronbach Alpha | .57 |
| Sentence Improvement-Onomatopoeia | Cronbach Alpha | .89 |
| Ethical Values Inventory: | | |
| Rational Conscientious | Cronbach Alpha | .76 |
| Irrational Conscientious | Cronbach Alpha | .44 |
| Group Conforming | Cronbach Alpha | .42 |
| Self-Centered | Cronbach Alpha | .98 |
| Handwriting Legibility | 2 independent raters | .94 |

## Treatment of the Data

The analysis of variance was used to determine the significance of differences among the seven groups in chronological age and IQ, criteria used in the selection process. Also, an analysis of variance for total battery achievement scores and for scores on the Teacher Rating Scale was run on the two second grade groups, originally matched by sex and then randomly assigned to acceleration or nonacceleration. When an $F$ significant at the .05 level was obtained, confidence intervals (Dunnett, 1955) were built around the differences between the accelerated group mean and each of the other six group means.[1]

---

[1] A table giving the mean scores and results of tests of significance for the 1960 data; a table giving the same information for the 1961 data; and a table showing the incidence of unusual physical, social, and emotional development have been deposited with the

A similar treatment was applied to all data obtained in March 1961. A two-way classification was used in the analysis of variance, sex by group, to test differences according to group, sex, and interactions between sex and group. When a significant interaction was found, confidence intervals were built separately for girls and boys. In addition, a chi square test was applied to the three development items on the Teacher Rating Scale. Chi square was used to test differences in the incidence of normal and unusual development of pupils in the seven groups.

When the confidence interval around a difference between means did not span zero, the difference was interpreted as significant. In Tables 2, 3, and 4 only the significance of differences is reported. A more detailed presentation of the confidence intervals is presented by Ripple (1961).

## Results

Differences between the accelerated groups and the six comparison groups on measures where the interaction was not significant are summarized in Table 2. The interpretations of these differences are based on the seven groups rather than on girls and boys separately. Since confidence intervals were built for the sexes separately on measures where a significant interaction was encountered, separate summary tables for girls and boys are presented for these measures, Table 3 for girls and Table 4 for boys.

Abbreviations are used to designate each of the seven groups of pupils in the tables as follows:

Acc—refers to the experimental group of pupils of superior learning abilities identified as older second graders in March 1960, and subsequently accelerated to the fourth grade in September 1960.

3SO—refers to the control group of superior learning abilities (the group randomly assigned to nonacceleration) identified as older second graders in March 1960 and subsequently enrolled in the third grade in September 1960.

3SY—refers to the control group of superior learning abilities identified as younger second graders in March 1960 and subsequently enrolled in the third grade in September 1960.

4SO—refers to the older control group of superior learning abilities identified on the basis of third grade (March 1960) IQ and achievement scores and subsequently enrolled in the fourth grade in September 1960.

4SY—refers to the younger control group of superior learning abilities

---

American Documentation Institute. Order Document No. 7073 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 2. Summary of Differences between the Accelerated Group and the Six Comparison Groups on Measures Where the Interaction Was Not Significant

| Measure | Not Significantly Different from Acc Group | Significantly Lower Than Acc Group | Significantly Higher Than Acc Group |
|---|---|---|---|
| March 1960 measures | | | |
| Chronological Age | 3SO | 3SY | 4SO, 4SY, 4AO, 4AY |
| IQ (Kuhlmann-Anderson) | 3SO, 4SO, 4SY | 4AO, 4AY | 3SY |
| Teacher Ratings | 3SO | — | — |
| Metropolitan Achievement Test-Elementary Battery | 3SO | — | — |
| March 1961 measures | | | |
| Metropolitan Achievement Test-Intermediate Battery: | | | |
| Arithmetic Computation | 4SY, 4AO | 3SO, 3SY, 4AY | 4SO |
| Arithmetic Problem Solving and Concepts | 4SY | 3SO, 3SY, 4AO, 4AY | 4SO |
| Attitude toward School, Learning: | | | |
| Places Liked Best | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Attitude Inventory | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Problem Solving Test | 3SO, 3SY, 4SY | 4AO, 4AY | 4SO |
| Ethical Values Inventory: | | | |
| Rational Conscientious | 3SO, 3SY, 4SO, 4SY, 4AY | 4AO | |
| Irrational Conscientious | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Group Conforming | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Handwriting Speed | 3SO, 3SY, 4SY, 4AO, 4AY | | 4SO |
| Psychomotor Abilities: | | | |
| Jump and Reach | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Wall Pass | 3SO, 3SY, 4SY, 4AO, 4AY | | 4SO |
| Shuttle Run | 3SO, 3SY, 4SO, 4AO, 4AY | | |
| Teacher Rating Scale: | | | |
| Total first 12 items | 3SO, 3SY, 4SO, 4SY | 4AO, 4AY | |
| Emotional Adjustment | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Social Adjustment | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Physical Coordination | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Creative Thinking Tests: | | | |
| Object Uses-Fluency | 3SO, 3SY, 4SY, 4AY | 4AO | 4SO |
| Object Uses-Flexibility | 3SO, 3SY, 4SY, 4AO, 4AY | | 4SO |
| Word Uses-Fluency | 3SO, 3SY, 4SY, 4AO, 4AY | | 4SO |
| Word Uses-Flexibility | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Plot Titles-Fluency | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Plot Titles-Cleverness | 3SY, 4SO | 3SO, 4SY, 4AO, 4AY | |
| Expressional Fluency | 3SO, 3SY, 4SO, 4SY | 4AO, 4AY | |
| Plot Questions Sensitivity | 4SY | 3SO, 3SY, 4AO, 4AY | 4SO |
| Object Improvement | 3SO, 3SY, 4SO, 4SY, 4AO | 4AY | |

identified on the basis of third grade (March 1960) IQ and achievement scores and subsequently enrolled in the fourth grade in September 1960.

4AO—refers to the older control group of average learning abilities identified on the basis of third grade (March 1960) IQ and achievement scores and subsequently enrolled in the fourth grade in September 1960.

4AY—refers to the younger control group of average learning abilities identified on the basis of third grade (March 1960) IQ and achievement scores and subsequently enrolled in the fourth grade in September 1960.

As can be seen in Table 2, the accelerated group (Acc) and the nonaccelerated control group (3SO) were not significantly different in chronological age, Kuhlmann-Anderson IQ, Metropolitan educational achievements, and teacher ratings used in the selection process in March 1960. Table 2 also presents differences between the accelerated group and the nonaccelerated control group (3SO) on measures obtained in March 1961, where the interaction between sex and group was not significant. It should be noted that the nonaccelerated control group was not significantly higher than the accelerated group on any of the measures in Table 2. On four measures—arithmetic computation, arithmetic problem solving and concepts, Plot Titles-Cleverness, and Plot Questions-Sensitivity—the accelerated group was significantly higher than the nonaccelerated control group.

Table 3 presents comparisons between the accelerated girls and girls in the six other groups on 11 measures where the interaction between sex and group was significant. It should be noted that the nonaccelerated control girls (3SO) were not significantly higher than the accelerated girls on any of the measures in Table 3. On four of the measures—total battery score, reading, spelling, and language total from the Metropolitan Achievement Tests—the accelerated girls were significantly higher than the nonaccelerated girls.

TABLE 3. Summary of Differences between Accelerated Girls and Girls in the Six Comparison Groups on 11 Measures Where the Interaction Was Significant

| Measure | Not Significantly Different from Acc Girls | Significantly Lower Than Acc Girls | Significantly Higher Than Acc Girls |
|---|---|---|---|
| March 1961 measures | | | |
| Metropolitan Achievement Test-Intermediate Battery: | | | |
| Total Battery Score | 4SY | 3SO, 3SY, 4AO, 4AY | 4SO |
| Word Knowledge | 3SO, 3SY, 4SY | 4AO, 4AY | 4SO |
| Reading | 4SO, 4SY | 3SO, 3SY, 4AO, 4AY | |
| Spelling | 3SY, 4SO, 4SY, 4AO | 3SO, 4AY | |
| Language Total | | 3SO, 3AY, 4AO, 4AY | 4SO, 4SY |
| Language Study Skills | 3SO, 4SO, 4SY, 4AO, 4AY | 3SY | |
| Social Studies Study Skills | 3SO, 4SO, 4SY, 4AY | 3SY, 4AO | |
| Handwriting Legibility | 3SO, 3SY, 4SO, 4SY, 4AY | 4AO | |
| Psychomotor Abilities: | | | |
| Basketball Throw | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Sociometric Instrument | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Creative Thinking Tests: | | | |
| Sentence Improvement | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |

TABLE 4. Summary of Differences between Accelerated Boys and Boys in the Six Comparison Groups on 11 Measures Where the Interaction Was Significant

| Measure | Not Significantly Different from Acc Boys | Significantly Lower Than Acc Boys | Significantly Higher Than Acc Boys |
|---|---|---|---|
| March 1961 measures | | | |
| Metropolitan Achievement Test-Intermediate Battery: | | | |
| Total Battery Score | 3SO, 3SY, 4SY | 4AO, 4AY | 4SO |
| Word Knowledge | 3SO, 3SY, 4SY | 4AO, 4AY | 4SO |
| Reading | 3SO, 3SY, 4SY | 4AO, 4AY | 4SO |
| Spelling | 3SO, 3SY, 4SO, 4SY, 4AY | 4AO | |
| Language Total | 3SO, 3SY, 4SY | 4AO, 4AY | 4SO |
| Language Study Skills | 3SO, 3SY, 4SO, 4SY | 4AO, 4AY | |
| Social Studies Study Skills | 3SO, 3SY, 4SO, 4SY | 4AO, 4AY | |
| Handwriting Legibility | 3SO, 3SY, 4SO, 4SY, 4AO, 4AY | | |
| Psychomotor Abilities: | | | |
| Basketball Throw | 3SO, 3SY, 4SY, 4AO, 4AY | | 4SO |
| Sociometric Instrument | 3SY, 4SY, 4AO, 4AY | | 3SO, 4SO |
| Creative Thinking Tests: | | | |
| Sentence Improvement | 3SO, 3SY, 4SY, 4AO, 4AY | | 4SO |

Table 4 presents comparisons between the accelerated boys and boys in the six other groups on 11 measures where the interaction between sex and group was significant. The nonaccelerated boys (3SO) were significantly higher than the accelerated boys only in acceptance by peers as measured with the sociometric instrument. These two groups of boys were not significantly different on the other 10 measures presented in Table 4.

The many measures on which the accelerated pupils were not significantly different from or higher than the nonaccelerated control pupils can be observed in Tables 2, 3, and 4. Also, the comparisons of the accelerated pupils with pupils in the other five groups can be observed in the tables. In general, the accelerated pupils (Acc) were significantly higher than the younger third graders of superior learning abilities (3SY), significantly lower than older fourth graders of superior learning abilities (4SO), not significantly different from younger fourth graders of superior learning abilities (4SY), and significantly higher than older and younger fourth graders of average learning abilities (4AO and 4AY).

## Discussion

Terman (Terman & Oden, 1947), Lehman (1953), and Pressey (1955), and many others have cited the need for early identification of superior pupils and have recommended that these pupils progress more rapidly through school and college. Since the accelerated pupils in this experiment were among the oldest for their grade, a greater sense of urgency surrounds the question of their acceleration. Assuming the usual progression of one grade per year, the accelerated pupils in this study will graduate from high school at age 17-1 to 17-7 rather than 18-1 to 18-7. This will permit college entrance and eventually starting their productive careers one year earlier than would have been possible without the acceleration.

The data on which the comparisons were based were collected in the seventh month of the first school year following the acceleration. Since the accelerates performed as well or better than their older third grade controls at that time, their acceleration was not harmful. The only observed negative effect was relatively lower peer acceptance for the accelerated boys. Subsequent research will determine if this decrease was only temporary. The accelerated and other pupils will be studied as they progress through their school careers.

To the extent that enabling students of superior learning abilities to graduate from high school at age 17 rather than age 18 is a worthwhile end, the results of this experiment are interpreted as strongly favorable toward accelerating older second graders of superior learning abilities to the fourth grade following a 5-week summer session.

## REFERENCES

Dunnett, C. W. A multiple comparison procedure for comparing several treatment means with a control. *J. Amer. Statist. Ass.*, 1955, **50**, 1096–1121.

Guilford, J. P., Kettner, N. W., & Christensen, P. R. A factor analytic study across the domains of reasoning, creativity, and evaluation: II. Administration of tests and analysis of results. *Univ. Sth. Calif. Lab. Rep.*, 1956, No. 16.

Klausmeier, H. J. Critical questions related to the education of the gifted. *Educ.*, 1958, **79**, 39–44.

Lehman, H. C. *Age and achievement*. Princeton: Princeton Univer. Press, 1953.

May, F. B. Creative thinking: A factorial study of seventh-grade children. Unpublished doctoral dissertation, Univer. of Wisconsin, 1961.

Pressey, S. L. Acceleration: Basic principles and recent research. In *Proceedings, international conference on testing problems*. Princeton: Educational Testing Service, 1955. Pp. 107–112.

Ripple, R. E. A controlled experiment in acceleration from the second to the fourth grade. Unpublished doctoral dissertation, Univer. of Wisconsin, 1961.

Terman, L. M., & Oden, Melita H. *The gifted child grows up: Twenty-five years' follow-up of a superior group*. Stanford: Stanford Univer. Press, 1947.

# 10

# Sex Differences in the Life Problems and Interests of Adolescents, 1935 and 1957

## — *Dale B. Harris*

*Among the many characteristics in which learners differ from one another are those relating to interests and personal problems. Variations in these should not be overlooked by educators when making decisions about the presentation of materials to be learned in school. In the study reported here, the interests and problems of adolescents in 1935 and 1957 are compared. Fifteen topics were ranked by adolescents in two ways: as personal problems and as items of interest. Note the differences in the rankings between girls and boys, between topics ranked as personal problems and as items of interest, and in the pattern of responses over the 22-year period.*

THIS IS a study of interests and problems which uses the method of rank order rather than a check list as its research technique. The items ranked are 15 topics selected from the concerns of adolescents. High school students were asked first to consider the items as personal problems and to construct an order reflecting their own experience with the issues as personal problems. The students then reranked the same items in order of interest, considering the topics as things they would like to read about and discuss or hear discussed.

Reprinted from *Child Development,* 1959, **30**, 453–459, by permission of the author and the Society for Research in Child Development.

The problems and instructions for ranking were taken verbatim from a study published by Symonds (*1*, *2*). He had selected the issues from young people's own discussions and phrased the issues in terms used by young people themselves. His 1641 students attended junior and senior high schools in Tulsa, Oklahoma, and New York City. The 1165 youth in the present study came from the junior-senior high school in a Minnesota community. Twenty-two years and considerable social, cultural, and economic change separate the circumstances of the two studies. Whether geographic or regional differences also influence the data cannot be known. In both studies, the samples represented in general the socioeconomic distribution of the communities from which they were drawn. The comparability of samples drawn from different geographic areas without close control through a stratification procedure is questionable. But the comparability of samples is also problematical across long time intervals, in which social and cultural changes have occurred, even when stratification by some socioeconomic index has been attempted. Changing conditions may themselves affect the index.

The hypothesis of the present study is taken from Symonds' discussion (*2*): "Change the social and economic structure of society and you immediately change the relative emphasis of these problems and interests" (p. 752).

In a ranking process as in any *system*, change in one feature or aspect may have widespread effects throughout the system. Thus, there are some limitations to the method. The elements are *relative* to one another. The resulting rank order is not a true *scale* of values. Each choice made removes a degree of freedom and relates to the choices already made and those yet to be made. The ranks accorded a series of stimuli are systematically affected by the order in which the stimuli are first presented. Symonds removed this effect in his study by presenting the items in reverse order to approximately half his subjects. This study used the same procedure. The sample of children was drawn randomly from the available supply to constitute groups of 100 boys and 100 girls at each grade; half of each sex responded to the items in the order Symonds presented them in his report, and the other half responded to a reversed order.

For 15 items ranked in one experience, a rho value of .535 is required to establish a relationship greater than chance at the 5 per cent level of significance. To interpret differences in rankings accorded items by boys and girls, or changes between two sets of rankings made at different times, one can place more confidence in *mean* ranks of samples than in ranks made by individuals. But the meaning of the magnitude of the differences thus described is still elusive. By building a body of experience with different arrangements of the data, one can possibly increase the insight he brings to bear on any one arrangement.

For example, when means of rankings of items considered as problems, established by 100 boys in each grade, 7 through 12, are compared, the average or typical between grade-group rank order correlation is +.74 (estimated from Kendall's W). For the same items considered as interests, the typical between grade-group rank order correlation is +.78. For girls, the corresponding values are +.70 and +.72. Thus, children's rankings from grade to grade change no more on the average than is suggested by correlation values in the .70's. Putting the data of boys and girls together in each grade, thus increasing reliability through the size of *N* but also introducing any existing systematic sex differences, the values become +.65 (problems) and +.63 (interests), (both estimated from Kendall's W).

Because there are likely to be intrinsic relationships in any set of values ranked by a group of judges, simply because judges from similar backgrounds are not reacting blindly or randomly to meaningful material, it is instructive to examine some typical interarray correlations estimated from Kendall's W. If we consider the four ranks of mean values accorded by boys and girls in 1935 and 1957 to *problems*, this typical interarray correlation is +.58. For four arrays of *interests* this value is also +.58. Throwing together the eight arrays of problems and interests, the value becomes +.36. These values provide a pragmatic bench-mark to judge the meaning of the correlation values presented below, which have a theoretical top value of +1.00 and describe a chance relationship as .00.

Taking all boys regardless of grade and similarly combining all girls for purposes of comparison with data obtained on the same items by Symonds in 1935, some interesting observations can be made. Youth showed greater consistency in the ordering of these items, considered first as problems and then as interests, in 1935 than they do at present. This change is true both for boys (+.63 compared with −.23) and for girls (+.66 compared with +.25). Thus, in 1957 there is little correspondence between a set of issues considered as problems and the order of interest in the same issues. This was not true in 1935, when there was a noticeable correspondence between the sets of ranks.

But how do young people separated by almost a generation compare in the ordering of their problems and their interests? Boys' rankings of interests across the years are somewhat more consistent (+.76) than their rankings of problems then and now (+.47). For girls, there is little difference between relative positions accorded problems and interests over the years (problems, +.50; interests, +.55).

Do boys and girls accord the same order of importance to a set of adolescent problems? In both periods boys and girls rank their problems similarly. In 1935 the similarity of the rank order of the sexes is expressed by a correlation value of +.80. In the 1957 study the value is +.77. The similarity of boys and girls is as great, on the average, as the similarity of successive grade groups of boys, or of girls. When the items were ranked according to interest in 1935, a value of +.80 expresses the similarity of boys' and girls' judgments. The comparable figure in 1957 is +.58.

The changes described above become more interesting when we look at specific ranks in Tables 1 and 2. As problems (Table 1), three items change five ranks or more across time for both boys and girls (health, mental health, manners), three more for boys only (safety, recreation, sched-

TABLE 1. Ranks Accorded Issues Considered As Problems by High School Boys and Girls in 1935 and in 1957

| | Boys | | | | Girls | | | |
|---|---|---|---|---|---|---|---|---|
| | 1935 | | 1957 | | 1935 | | 1957 | |
| Issue | Mean Rank | Rank | Rank | Mean Rank | Mean Rank | Rank | Mean Rank | Rank |
| Health | 6.7 | (2) | 9.1° | (12) | 6.6 | (2) | 8.7° | (12.5) |
| Love, marriage | 10.8 | (15) | 9.2° | (13.5) | 11.0 | (15) | 8.5° | (10) |
| Safety | 8.3 | (8.5) | 9.2° | (13.5) | 8.8 | (12) | 10.0° | (14.5) |
| Money | 6.2 | (1) | 6.3 | (2) | 6.8 | (3) | 6.5 | (2) |
| Mental hygiene | 8.7 | (13) | 8.2 | (8) | 8.2 | (9.5) | 6.9° | (3.5) |
| Study habits | 6.8 | (3) | 5.0 | (1) | 7.4 | (6) | 6.3° | (1) |
| Recreation | 8.3 | (8.5) | 10.1 | (15) | 8.4 | (11) | 10.0° | (14.5) |
| Personal, moral qualities | 7.1 | (4) | 6.6° | (3) | 7.3 | (4) | 7.1 | (5) |
| Home, family relationships | 8.2 | (7) | 8.4 | (10) | 8.2 | (9.5) | 7.6 | (6) |
| Manners | 8.5 | (11.5) | 7.6° | (5) | 7.4 | (6) | 8.6° | (11) |
| Personal attractiveness | 7.9 | (6) | 7.7 | (6) | 6.2 | (1) | 6.9° | (3.5) |
| Daily schedule | 9.0 | (14) | 8.3° | (9) | 9.4 | (14) | 8.7° | (12.5) |
| Civic interest | 8.5 | (11.5) | 7.9° | (7) | 9.0 | (13) | 8.4 | (9) |
| Getting along with other people | 8.4 | (10) | 8.6 | (11) | 7.9 | (8) | 8.0 | (8) |
| Philosophy of life | 7.5 | (5) | 7.5 | (4) | 7.4 | (6) | 7.7 | (7) |

° Change from 1935 significant at the 1 per cent level.

TABLE 2. Ranks Accorded Issues Considered As Interests by High School Boys and Girls in 1935 and in 1957

| | Boys | | | | Girls | | | |
|---|---|---|---|---|---|---|---|---|
| | 1935 | | 1957 | | 1935 | | 1957 | |
| Issue | Mean Rank | Rank | Rank | Mean Rank | Mean Rank | Rank | Mean Rank | Rank |
| Health | 5.6 | (2) | 6.4° | (2) | 6.6 | (4) | 7.0 | (6) |
| Love, marriage | 9.3 | (13) | 7.7° | (8) | 9.4 | (12) | 6.5° | (4.5) |
| Safety | 7.8 | (7) | 8.7° | (11) | 9.2 | (10) | 10.2° | (14) |
| Money | 7.1 | (3) | 6.5° | (3) | 8.1 | (8) | 8.2 | (10) |
| Mental hygiene | 9.6 | (14) | 9.5 | (13) | 9.8 | (13.5) | 8.0° | (9) |
| Study habits | 8.7 | (11) | 9.4° | (12) | 9.3 | (11) | 9.9° | (12.5) |
| Recreation | 4.9 | (1) | 5.6° | (1) | 5.6 | (2) | 7.8° | (8) |
| Personal, moral qualities | 7.7 | (5.5) | 7.3° | (6) | 7.6 | (7) | 7.1 | (7) |
| Home, family relationships | 8.4 | (10) | 7.2° | (5) | 8.3 | (9) | 6.4° | (2.5) |
| Manners | 7.5 | (4) | 8.6° | (10) | 6.3 | (3) | 8.5° | (11) |
| Personal attractiveness | 8.1 | (8) | 8.0 | (9) | 5.4 | (1) | 6.0° | (1) |
| Daily schedule | 10.5 | (15) | 10.9 | (15) | 10.4 | (15) | 11.4° | (15) |
| Civic interest | 9.0 | (12) | 9.6° | (14) | 9.8 | (13.5) | 9.9 | (12.5) |
| Getting along with other people | 8.2 | (9) | 7.1° | (4) | 7.0 | (5) | 6.4° | (2.5) |
| Philosophy of life | 7.7 | (5.5) | 7.4 | (7) | 7.3 | (6) | 6.5° | (4.5) |

° Change from 1935 significant at the 1 per cent level.

ule), and two more for girls only (love and marriage, and study habits). As interests (Table 2), three items change for both sexes across time (love and marriage, family relations, manners), and one more for each sex singly (recreation for girls, getting along with others for boys). There are as many sex differences now as in 1935. At that time seven topics significantly differentiated the sexes as problems at the .01 level of certainty.[1] Now eight topics satisfy the .01 level.

Sex differences in interests also are about the same. In 1935 boys and girls rated nine topics quite different. In 1957, 10 topics satisfied the .01 level.

Attention to item placements show that a number of topics are relatively high as sources of problems: *Money* is still high as a problem to both boys and girls (ranks 2 and 2, respectively). It is of considerably greater interest to boys than to girls (rank 3 as compared with rank 10).[2] *Health* at rank 2 as a problem in 1935 is no longer seen as such (rank 12), though interest in it is still relatively high, especially among boys. *Study habits* are somewhat more of a problem now than in 1935, especially for boys, but the topic ranked quite low in interest value in both periods. *Moral qualities* and *philosophy of life* come next in position both as problems and as topics of interest and are of similar relative magnitude in both periods.

Of intermediate concern as problems are the following: Both sexes see *mental health* as somewhat more of a problem to them now than they did in 1935, and for girls it now appears in the top five ranks. It has, likewise, moved up significantly ($p = .01$) in rank of interest to girls, though it remains low on the boys' list. *Home and family relations* is likewise ranked higher as a problem by girls than by boys now ($p = .01$). Relative to 1935, boys rank it slightly lower and girls slightly higher as a problem, though in these intermediate positions such shifting is statistically meaningless. Both boys and girls rank this topic higher on the list of interests than they did in 1935, girls very considerably so (rank of 2.5 compared with 9, $p = .01$). *Manners and courtesy*, up somewhat ($p = .01$) as a problem for boys as compared with the earlier period, is significantly less a problem for girls now (rank 11 compared with rank 6, $p = .01$). As a topic of interest it is down sharply for both sexes. *Attractiveness*, of intermediate value both as a problem and as an interest for boys in both periods, ranks high in both interest value and as a problem for girls in both periods. Likewise, *getting along with others* was and is of considerable interest to girls and has risen (rank 4 compared to rank 9, $p = .01$) as a topic of interest to boys. It is of intermediate significance as a problem. *Civic affairs* was and is of about median significance as a problem to both boys and girls and is of even less relative importance as a topic of interest.

---

[1] This index of statistical significance refers to the difference between mean ranks, interpreted in terms of the standard errors of these means. It does not refer to shift in rank order.

[2] In 1935 Symonds attributed the high ranks accorded money to the current economic stress. Studies of allowances show that modern youth "never had it so good" from an economic point of view, but inflation and rising standards of living expectancy keep the 'teen-age group keenly aware of the medium of exchange.

Of least concern as problems now are the following: *Recreation*, of intermediate significance as a problem to both boys and girls in 1935, is negligible now. It was and is number 1 in interest value to boys but has dropped from second to eighth place for girls ($p = .01$). *Health*, of high significance both as a problem and as a topic of interest to boys and girls in 1935, is now negligible as a problem to either sex. It is now only of intermediate interest to girls, though it remains high on the boys' list. *Safety* is of little interest and even less a problem to both boys and girls now than in 1935 ($p = .01$). *Love and marriage*, ranked low as a problem by boys in both periods, has risen to an intermediate rank as a problem for girls. Both sexes now give it an intermediate interest rating, some several ranks higher than 22 years ago. The change in both problem and interest values of this topic is statistically significant, highly so for girls. *Daily schedule* is of little significance then or now, both to boys and girls, as problem or as interest.

This observation concerning the specific rankings accorded the issues should be made. Both psychologically and statistically, the highest and lowest ranks are most differentiated in any ranking or scaling procedure. The method in this study called for the identification of the first three positions, and then positions 13, 14, and 15. The intermediate positions were assigned last and may, particularly in the ranking of interests, represent a state of psychological indifference or lack of discrimination more than a state of intermediate significance. Should this be the case, the significance attributed by Symonds to the "relatively high" rank of philosophy of life (6th in 15 issues) may be modified somewhat. He affirmed (*1*) that "values and goals are craved" by youth and challenged teacher and counselors by the rank accorded this topic. Symonds attributed the high rank of money as a problem to the depression years and observed "it is a pity" that money drops to a much lower rank in interest value. The problem significance of money was not just a function of depression years, it is now clear.

The lack of concern with love and marriage, identified by Symonds as "sex adjustments," puzzled him, and he explained the low ranks by reference to "repression." His own hypothesis concerning social change appears from the present data to be equally plausible. A similar point can be made about mental health as an issue in adolescence. Symonds dismissed the low ranks attributed to this issue in these words: "Mental health likewise is no concern of healthy, growing adolescents. The crest of life is before them. Their failures and thwartings have not yet turned them back upon themselves" (1, p. 517). The changed cultural ethos apparently has made a difference in the significance accorded this issue in the adolescent years.

The shifts noted in the tables and in the brief discussion for the most part confirm what observers of recent social trends have noted. Today, youth marry younger and show an earlier interest in social relations, love, and marriage. Our culture appears to recognize more openly now than two decades ago the sex, love, and marriage problems of young people. Physical health is actually less a problem today, and possibly receives less attention in school and in the popular press, whereas mental health discussions, literature, and posters appear in every newspaper, magazine, and waiting room. An increase in informality and casualness in dress and behavior may reflect itself in the decline in concern with manners.

The student of adolescent behavior will not be surprised at the significance of money as a problem, high interest in recreation, lack of concern with and interest in safety, unconcern over daily schedule and civic affairs, considerable concern over study habits as a problem but lack of interest in the topic, nor will he be surprised by the greater interest of girls than boys in attractiveness, love and marriage, mental health, and philosophy and beliefs; of boys than girls in money, health, and recreation.

If adults wish to "view with alarm," they may attend to the adolescent's relative unconcern with safety and hazard, set over against the 'teen-age driving problem, and the young person's continuing unconcern with civic affairs, set over against the continued increased emphasis on "modern problems," and citizenship in the secondary school's curriculum theory and effort.

## REFERENCES

1. Symonds, P. M. Life interests and problems of adolescents. *Sch. Rev.*, 1936, 44, 506–518.
2. Symonds, P. M. Sex differences in the life problems and interests of adolescents. *Sch. & Soc.*, 1936, 43, 751–752.

# 11

# The Adolescent Subculture and Academic Achievement — *James S. Coleman*

*Variations in the value systems of learners are strongly dependent upon environmental factors such as the home, the school, and the church. In this article the author examines the sources of adolescent value systems. He reports research on the hypothesis that, in an adolescent social system in which academic achievement is greatly valued, the group with highest achievement will include more students whose actual intelligence is high than will a similar group from a social system where academic achievement is less valued. The effect of interscholastic athletics on adolescent value systems is considered. The provocative suggestion that interscholastic academic competition should be encouraged in order to knit intraschool cohesion and to help increase the value of academic achievement should prove a stimulus to further discussion.*

INDUSTRIAL SOCIETY has spawned a peculiar phenomenon, most evident in America but emerging also in other Western societies: adolescent subcultures, with values and activities quite distinct from those of the adult society—subcultures whose members have most of their important associations within and few with adult society. Industrialization, and the rapidity of change itself, has taken out of the hands of the parent the task of training his child, made the parent's skills obsolescent, and put him out of touch with the times—unable to understand, much less inculcate, the standards of a social order which has changed since he was young.

Reprinted from *American Journal of Sociology*, 1960, **65**, 337–347, by permission of The University of Chicago Press. © The University of Chicago Press.

By extending the period of training necessary for a child and by encompassing nearly the whole population, industrial society has made of high school a social system of adolescents. It includes, in the United States, almost all adolescents and more and more of the activities of the adolescent himself. A typical example is provided by an excerpt from a high-school newspaper in an upper-middle-class suburban school:

> SOPHOMORE DANCING FEATURES CHA CHA
>
> SOPHOMORES, this is your chance to learn how to dance! The first day of sophomore dancing is Nov. 14 and it will begin at 8:30 A.M. in the Boys' Gym. . . .
>
> NO ONE is required to take dancing but it is highly recommended for both boys and girls. . . .
>
> If you don't attend at this time except in case of absence from school, you may not attend at any other time. Absence excuses should be shown to Miss ---- or Mr. ----.

In effect, then, what our society has done is to set apart, in an institution of their own, adolescents for whom home is little more than a dormitory and whose world is made up of activities peculiar to their fellows. They have been given as well many of the instruments which can make them a functioning community: cars, freedom in dating, continual contact with the opposite sex, money, and entertainment, like popular music and movies, designed especially for them. The international spread of "rock-and-roll" and of so-called American patterns of adolescent behavior is a consequence, I would suggest, of these economic changes which have set adolescents off in a world of their own.

Yet the fact that such a subsystem has sprung up in society has not been systematically recognized in the organization of secondary education. The theory and practice of education remains focused on *individuals;* teachers exhort individuals to concentrate their energies in scholarly directions, while the community of adolescents diverts these energies into other channels. The premise of the present research is that, if educational goals are to be realized in modern society, a fundamentally different approach to secondary education is necessary. Adults are in control of the institutions they have established for secondary education; traditionally, these institutions have been used to mold children as individuals toward ends which adults dictate. The fundamental change which must occur is to shift the focus: to mold social communities as communities, so that the norms of the communities themselves reinforce educational goals rather than inhibit them, as is at present the case.

The research being reported is an attempt to examine the status systems of the adolescent communities in ten high schools and to see the effects of these status systems upon the individuals within them. The ten high schools are all in the Midwest. They include five schools in small towns (labeled *0–4* in the figures which follow), one in a working-class suburb (*6*), one in a well-to-do suburb (*9*), and three schools in cities of varying sizes (*5*, *7*, and *8*). All but No. 5, a Catholic boys' school, are coeducational, and all but it are public schools.

The intention was to study schools which had quite different status systems, but the similarities were far more striking than the differences. In a

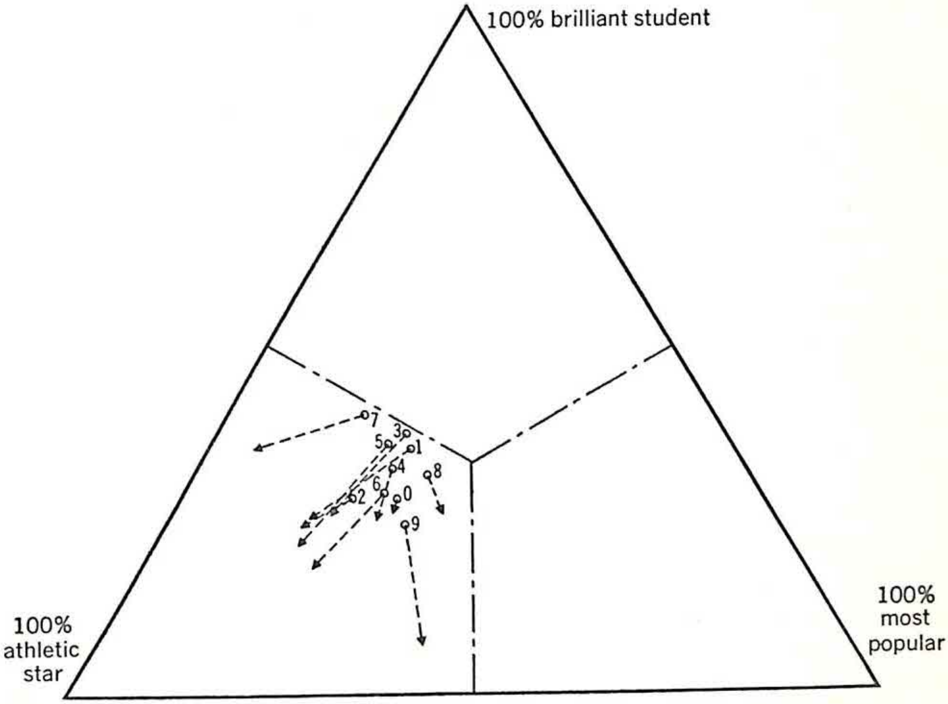


FIG. 1. *Positions of schools and leading crowds in boys' relative choice of brilliant student, athletic star, and most popular.*

questionnaire all boys were asked: "How would you most like to be remembered in school: as an athletic star, a brilliant student, or most popular? The results of the responses for each school are shown in Figure 1,[1] where the left corner of the triangle represents 100 per cent saying "star athlete"; the top corner represents 100 per cent saying "brilliant student"; and the right corner represents 100 per cent saying "most popular." Each school is representedly a point whose location relative to the three corners shows the proportion giving each response.

The schools are remarkably grouped somewhat off-center, showing a greater tendency to say "star athlete" than either of the other choices. From each school's point is a broken arrow connecting the school as a whole with its members who were named by their fellows as being "members of the leading crowd." In almost every case, the leading crowd tends in the direction of the athlete—in all cases *away* from the ideal of the brilliant student. Again, for the leading crowds as well as for the students as a whole, the uniformity is remarkably great; not so great in the absolute positions of the leading crowds but in the direction they deviate from the student bodies.

---

[1] I am grateful to James A. Davis and Jacob Feldman, of the University of Chicago, for suggesting such graphs for presenting responses to trichotomous items in a population.

This trend toward the ideal of the athletic star on the part of the leading crowds is due in part to the fact that the leading crowds include a great number of athletes. Boys were asked in a questionnaire to name the best athlete in their grade, the best student, and the boy most popular with girls. In every school, without exception, the boys named as best athletes were named more often—on the average over twice as often—as members of the leading crowd than were those named as best students. Similarly, the boy most popular with girls was named as belonging to the leading crowd more often than the best student, though in all schools but the well-to-do suburb and the smallest rural town (schools 9 and 0 on Fig. 1) less often than the best athlete.

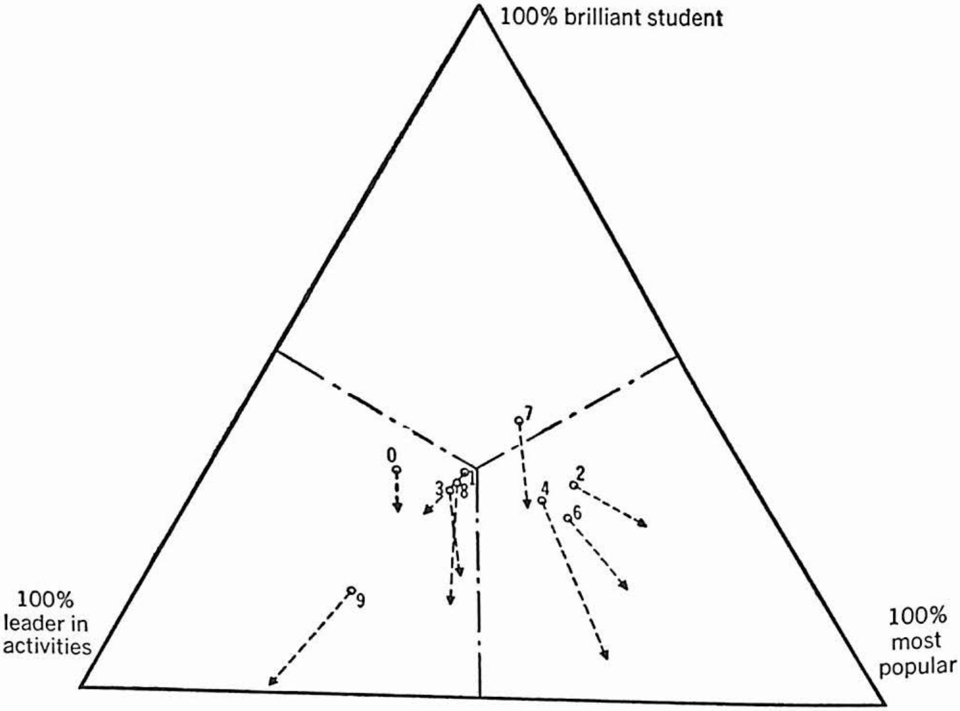


FIG. 2. *Title as in Fig. 6, but without last parentheses.*

These and other data indicate the importance of athletic achievement as an avenue for gaining status in the schools. Indeed, in the predominantly middle-class schools, it is by far the most effective achievement for gaining a working-class boy entrée into the leading crowd.

Similarly, each girl was asked how she would like to be remembered: as a brilliant student, a leader in extracurricular activities, or most popular. The various schools are located on Figure 2, together with arrows connecting them to their leading crowd. The girls tend slightly less, on the average, than the boys to want to be remembered as brilliant students. Although the alternatives are different, and thus cannot be directly compared, a great deal of other evidence indicates that the girls—although better students in every school—do not want to be considered "brilliant students." They have good reason not to, for the girl in each grade in each of the schools who was most often named as best student has fewer friends and is less often in the leading crowd than is the boy most often named as best student.

There is, however, diversity among the schools in the attractiveness of the images of "activities leader" and "popular girl" (Fig. 2). In five (9, 0, 3, 8, and 1), the leader in activities is more often chosen as an ideal than is the popular girl; in four (7, 6, 2, and 4) the most popular girl is the more attractive of the two. These differences correspond somewhat to class background differences among the schools: 2, 4, 6, and 7, where the activities leader is least attractive, have the highest proportion of students with working-class backgrounds. School 9 is by far the most upper-middle-class one and by far the most activities-oriented.

The differences among the schools correspond as well to differences among the leading crowds: in schools 2, 4, and 6, where the girls as a whole are most oriented to being popular, the leading crowds are even more so; in the school where the girls are most oriented to the ideal of the activities leader, No. 9, the leading crowd goes even further in that direction.[2] In other words, it is as if a pull is exerted by the leading crowd, bringing the rest of the students toward one or the other of the polar extremes. In all cases, the leading crowd pulls away from the brilliant-student ideal.

Although these schools vary far less than one might wish when examining the effects of status systems, there are differences. All students were asked in a questionnaire: "What does it take to get into the leading crowd?" On the basis of the answers, the relative importance of various activities can be determined. Consider only a single activity, academic achievement. Its importance for status among the adolescents in each school can be measured simply by the proportion of responses which specify "good grades," or "brains" as adolescents often put it, as a means of entrée into the leading crowd. In all the schools, academic achievement was of less importance than other matters, such as being an athletic star among the boys, being a cheerleader or being good-looking among the girls, or other attributes. Other measures which were obtained of the importance of academic achievement in the adolescent status system correlate highly with this one.[3]

---

[2] This result could logically be a statistical artifact because the leaders were included among students as a whole and thus would boost the result in the direction they tend. However, it is not a statistical artifact, for the leading crowds are a small part of the total student body. When they are taken out for computing the position of the rest of the girls in each school, schools 2, 4, 6, and 7 are still the most popularity-oriented, and school 9 the most activities-oriented.

If, then, it is true that the status system of adolescents *does* affect educational goals, those schools which differ in the importance of academic achievement in the adolescent status system should differ in numerous other ways which are directly related to educational goals. Only one of those, which illustrates well the differing pressures upon students in the various schools, will be reported here.

In every social context certain activities are highly rewarded, while others are not. Those activities which are rewarded are the activities for which there is strong competition—activities in which everyone with some ability will compete. In such activities the persons who achieve most should be those with most potential ability. In contrast, in unrewarded activities, those who have most ability may not be motivated to compete; consequently, the persons who achieve most will be persons of lesser ability. Thus in a high school where basketball is important, nearly every boy who might be a good basketball player will go out for the sport, and, as a result, basketball stars are likely to be the boys with the most ability. If in the same school volleyball does not bring the same status, few boys will go out for it, and those who end up as members of the team will not be the boys with most potential ability.

Similarly, with academic achievement: in a school where such achievement brings few social rewards, those who "go out" for scholarly achievement will be few. The high performers, those who receive good grades, will not be the boys whose ability is greatest but a more mediocre few. Thus the "intellectuals" of such a society, those defined by themselves and others as the best students, will not in fact be those with most intellectual ability. The latter, knowing where the social rewards lie, will be off cultivating other fields which bring social rewards.

To examine the effect of varying social pressures in the schools, academic achievement, as measured by grades in school, was related to I.Q. Since the I.Q. tests differ from school to school, and since each school had its own mean I.Q. and its own variation around it, the ability of high performers

---

[3] Parenthetically, it might be noted that these measures correlate imperfectly with the proportion of boys or girls who want to be remembered as brilliant students. These responses depend on the relative attractiveness of other ideals, which varies from school to school, and upon other factors unrelated to the status system.

(boys who made *A* or *A*–, average)[4] was measured by the number of standard deviations of their average I.Q.'s above the mean. In this way, it is possible to see where the high performers' ability lay, relative to the distribution of abilities in their school.[5]

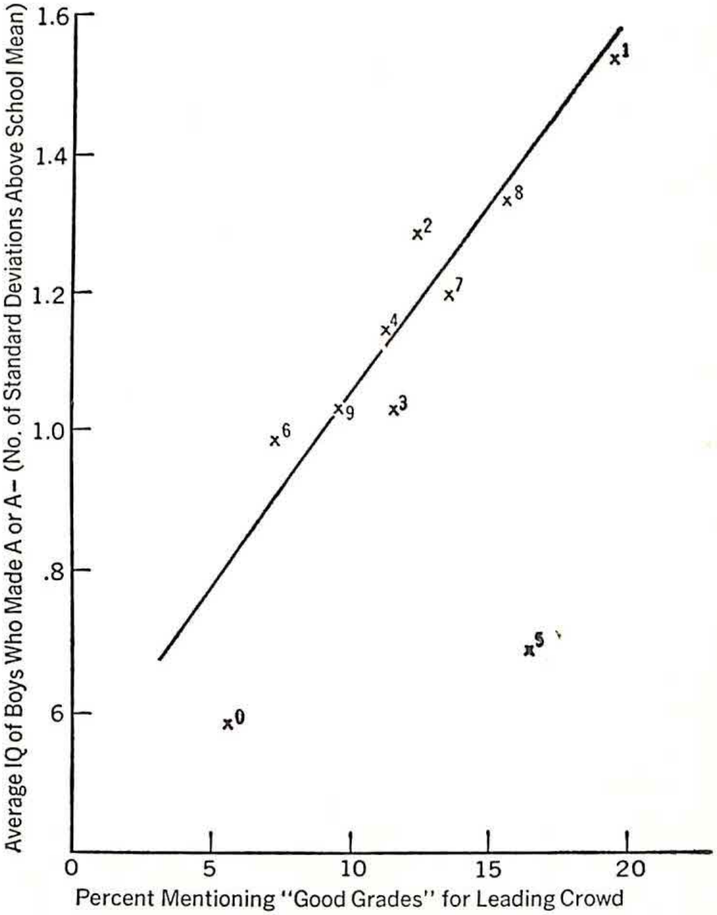


FIG. 3. *I.Q.'s of high achieving boys by importance of good grades among other boys.*

---

[4] In each school but 3 and 8, those making *A* and *A*– constituted from 6 to 8 per cent of the student body. In order to provide a correct test of the hypothesis, it is necessary to have the same fraction of the student body in each case (since I.Q.'s of this group are being measured in terms of number of standard deviations above the student body). To adjust these groups, enough 6's were added (each being assigned the average I.Q. of the total group of 6's) to bring the proportion up to 6 per cent (from 3 per cent in school 3, from 4 per cent in school 8).

[5] The I.Q. tests used in the different schools were: (0) California Mental Maturity (taken seventh, eighth, or ninth grade); (1) California Mental Maturity (taken eighth grade); (2) SRA Primary Mental Abilities (taken tenth grade); (3) California Mental Maturity (taken ninth grade; seniors took SRA PMA, which was tabulated as a percentile, and they have been omitted from analysis reported above); (4) Otis (ninth and

The variations were great: in a small-town school, No. *1*, the boys who made an *A* or *A*− average had I.Q.'s 1.53 standard deviations above the school average; in another small-town school, No. *0*, their I.Q.'s were only about a third this distance above the mean, .59. Given this variation, the question can be asked: Do these variations in ability of the high performers correspond to variations in the social rewards for, or constraints against, being a good student?

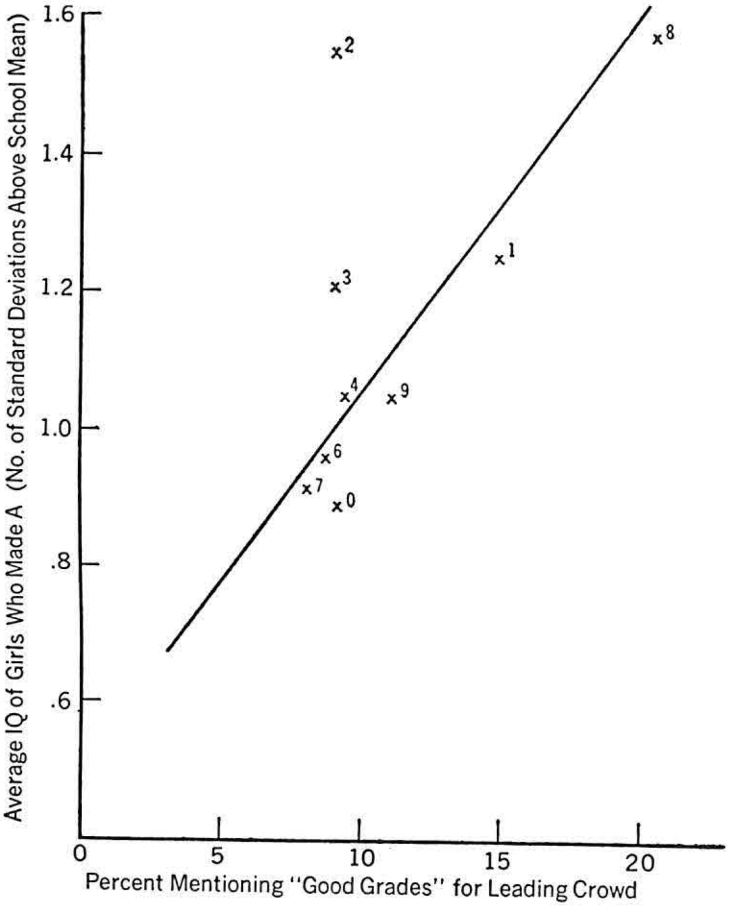


FIG. 4. *I.Q.'s of high achieving girls by importance of good grades among other girls.*

Figure 3 shows the relation for the boys between the social rewards for academic excellence (i.e., the frequency with which "good grades" was mentioned as a means for getting into the leading crowd) and the ability of the

---

tenth grades; taken eighth grade); Kuhlman Finch (eleventh and twelfth grades, taken eighth grade); (*5*) Otis (taken ninth grade); (*6*) California Mental Maturity (taken eighth grade); (*7*) California Mental Maturity (taken eighth grade); (*8*) Otis (taken ninth or tenth grade); and (*9*) Otis (taken eighth grade).

high performers; measured by the number of standard deviations their average I.Q.'s exceed that of the rest of the boys in the school. The relation is extremely strong. Only one school, a parochial boys' school in the city's slums, deviates. This is a school in which many boys had their most important associations outside the school rather than in it, so that its student body constituted far less of a social system, less able to dispense social rewards and punishments, than was true of the other schools.

Similarly, Figure 4 shows for the girls the I.Q.'s of the high performers.[6] Unfortunately, most of the schools are closely bunched in the degree to which good grades are important among the girls, so that there is too little variation among them to examine this effect as fully as would be desirable. School 2 is the one school whose girls deviate from the general relationship.

The effect of these values systems on the freedom for academic ability to express itself in high achievement is evident among the girls as it is among the boys. This is not merely due to the school facilities, social composition of the school, or other variables: the two schools highest in the importance of scholastic achievement for both boys and girls are *1* and 8, the first a small-town school of 350 students and the second a city school of 2,000 students. In both there are fewer students with white-collar backgrounds than in schools 9 or 3, which are somewhere in the middle as to value placed on academic achievement, but are more white-collar than in schools *7* or *4*, which are also somewhere in the middle. The highest expenditure per student was $695 per year in school 9, and the lowest was little more than half that, in school *4*. These schools are close together on the graphs of Figures 3 and 4.

It should be mentioned in passing that an extensive unpublished study throughout Connecticut, using standard tests of achievement and ability, yielded consistent results. The study found no correlation between per pupil expenditure in a school and the achievement of its students relative to their ability. The effects shown in Figures 3 and 4 suggest why: that students with ability are led to achieve only when there are social rewards, primarily from their peers, for doing so—and these social rewards seem little correlated with per pupil expenditure.

So much for the effects as shown by the variation among schools. As mentioned earlier, the variation among schools was not nearly so striking in this research as the fact that, in all of them, academic achievement did not count for as much as other activities. In every school the boy named as best athlete and the boy named as most popular with girls was far more often mentioned as a member of the leading crowd, and as someone to "be like," than was the boy named as the best student. And the girl named as best dressed, and the one named as most popular with boys, was in every school far more often mentioned as being in the leading crowd and as someone "to be like," than was the girl named as the best student.

---

[6] For the girls, only girls with a straight-*A* average were included. Since girls get better grades than boys, this device is necessary in order to make the sizes of the "high-performer" group roughly comparable for boys and for girls. Schools differed somewhat in the proportion of *A*'s, constituting about 6 per cent of the students in the small schools, only about 3 per cent in schools 6 and *7*, 1 per cent in 8, and 2 per cent in 9. In 8 and 9, enough girls were added and assigned the average grade of the *7* (*A*–) group to bring the proportion to 3 per cent, comparable with the other large schools. The difference, however, between the large and small schools was left.

The relative unimportance of academic achievement, together with the effect shown earlier, suggests that these adolescent subcultures are generally deterrents to academic achievement. In other words, in these societies of adolescents those who come to be seen as the "intellectuals" and who come to think so of themselves are not really those of highest intelligence but are only the ones who are willing to work hard at a relatively unrewarded activity.

The implications for American society as a whole are clear. Because high schools allow the adolescent subcultures to divert energies into athletics, social activities, and the like, they recruit into adult intellectual activities people with a rather mediocre level of ability. In fact, the high school seems to do more than allow these subcultures to discourage academic achievement; it aids them in doing so. To indicate how it does and to indicate how it might do differently is another story, to be examined below.

Figures 1 and 2, which show the way boys and girls would like to be remembered in their high school, demonstrate a curious difference between the boys and the girls. Despite great variation in social background, in size of school (from 180 to 2,000), in size of town (from less than a thousand to over a million), and in style of life of their parents, the proportion of boys choosing each of the three images by which he wants to be remembered is very nearly the same in all schools. And in every school the leading crowd "pulls" in similar directions: at least partly toward the ideal of the star athlete. Yet the ideals of the girls in these schools are far more dispersed, and the leading crowds "pull" in varying directions, far less uniformly than among the boys. Why such a diversity in the same schools?

The question can best be answered by indirection. In two schools apart from those in the research, the questionnaire was administered primarily to answer a puzzling question: Why was academic achievement of so little importance among the adolescents in school *9*? Their parents were professionals and business executives, about 80 per cent were going to college (over twice as high a proportion as in any of the other schools), and yet academic excellence counted for little among them. In the two additional schools parental background was largely held constant, for they were private, coeducational day schools whose students had upper-middle-class backgrounds quite similar to those of school *9*. One (No. *10*) was in the city; the other (No. *11*), in a suburban setting almost identical to that of No. 9. Although the two schools were added to the study to answer the question about school 9, they will be used to help answer the puzzle set earlier: that of the clustering of schools for the boys and the greater spread for the girls. When we look at the responses of adolescents in these two schools to the question as to how they would like to be remembered, the picture becomes even more puzzling (Figs. 5 and 6). For the boys, they are extremely far from the cluster of the other schools; for the girls, they are intermingled with the other schools. Thus, though it was for the boys that the other schools clustered so closely, these two deviate sharply from the cluster; and for the girls, where the schools already varied, these two are not distinguishable. Furthermore, the leading crowds of boys in these schools do not pull the ideal toward the star-athlete ideal as do those in almost all the other schools. To be sure, they pull away from the ideal of the brilliant student, but the pull is primarily toward a social image, the most popular. Among the girls, the leading crowds pull in different directions and are nearly indistinguishable from the other schools.

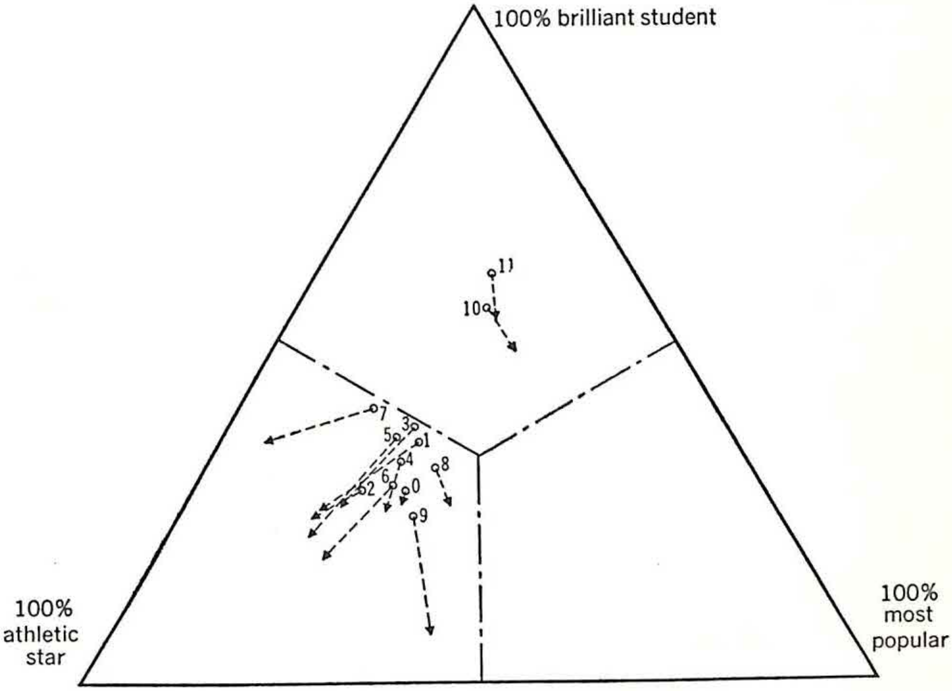


Fig. 5. *Title as in Fig. 1, but with added: (two private schools [10, 11] included).*

The answer to both puzzles, that is, first, the great cluster of the boys and now, in these two additional schools, the greater deviation, seems to lie in one fact: the boys' interscholastic athletics. The nine public schools are all engaged in interscholastic leagues which themselves are knit together in state tournaments. The other school of the first ten, the Catholic school, is in a parochial league, where games are just as hotly contested as in the public leagues and is also knit together with them in tournaments.

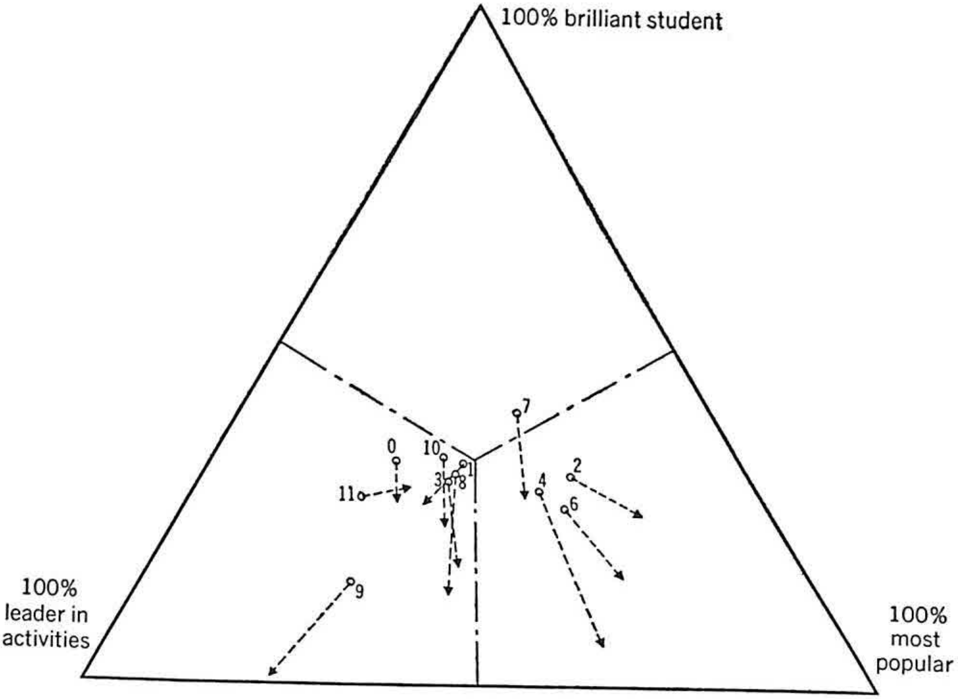


FIG. 6. *Positions of schools and leading crowds in girls' relative choice of brilliant student, activities leader, and most popular (two private schools* [10, 11] *included).*

Schools *10* and *11* are athletically in a world apart from this. Although boys in both schools may go in for sports, and both schools have interscholastic games, the opponents are scattered private schools, constituting a league in name only. They take no part in state or city tournaments and have almost no publicity.

There is nothing for the girls comparable to the boys' interscholastic athletics. There are school activities of one sort or another, in which most girls take part, but no interscholastic games involving them. Their absence and the lack of leagues which knit all schools together in systematic competition means that the status system can "wander" freely, depending on local conditions in the school. In athletics, however, a school, and the community surrounding it, cannot hold its head up if it continues to lose games. It *must* devote roughly the same attention to athletics as do the schools surrounding it, for athletic games are the only games in which it engages other schools and, by representation, other communities.

These games are almost the only means a school has of generating internal cohesion and identification, for they constitute the only activity in which the school participates as a school. (This is well indicated by the fact that a number of students in school *10*, the private school which engages in no interscholastic games, has been concerned by a "lack of school spirit.") It is as a consequence of this that the athlete gains so much status; he is doing something for the school and the community, not only for himself, in leading his team to victory, for it is a school victory.

The outstanding student, in contrast, has little or no way to bring glory to his school. His victories are always purely personal, often at the expense of his classmates, who are forced to work harder to keep up with him. It is no wonder that his accomplishments gain little reward and are often met by ridiculing remarks, such as "curve-raiser" or "grind," terms of disapprobation which have no analogues in athletics.

These results are particularly intriguing, for they suggest ways in which rather straightforward social theory could be used in organizing the activities of high schools in such a way that their adolescent subcultures would encourage, rather than discourage, the channeling of energies into directions of learning. One might speculate on the possible effects of city-wide or state-wide "scholastic fairs" composed of academic games and tournaments between schools and school exhibits to be judged. It could be that the mere institution of such games would, just as do the state basketball tournaments in the midwestern United States, have a profound effect upon the educational climate in the participating schools. In fact, by an extension of this analysis, one would predict that an international fair of this sort, a "Scholastic Olympics," would generate interscholastic games and tournaments within the participating countries.

# Part III

## *Teacher Characteristics and Classroom Interactions*

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

AMONG THOSE factors affecting the efficiency of learning are a teacher's personal characteristics and classroom interactions between the teacher and students. These interactions (including what the teacher and students say, do, and feel in the learning situation) largely determine the emotional setting of learning as well as the efficiency with which its goals are achieved. Seven selections pertaining to teacher characteristics and classroom interactions are presented in Part III.

An investigation based on a correlational approach of the interdependency of the pupil behavior-teacher behavior relationship is reported by David Ryans in the first selection. Ryans, a professor at the University of Texas at the time this article was published, is currently associated with the System Development Corporation and U. C. L. A. His analyses support the conclusion that teacher behavior and pupil behavior show substantially more interdependence in elementary school than they do in secondary school.

Kenneth Hoyt, professor of counseling and guidance in the College of Education at the University of Iowa, reports in the next selection an investigation on the belief that better teaching results from supplying teachers with pertinent facts about their pupils. More specifically, his research was concerned with studying the effects of teacher knowledge of pupil characteristics on pupil achievement and teacher-pupil relationships. Though pupil achievement was not affected, increases in teacher knowledge of pupil characteristics did improve pupil attitudes toward teachers.

In the next paper Harry Beilin, of Brooklyn College, reviews the thirty years of research following the publication of the Wickman study in 1928, which contrasted the attitudes of teachers and clinicians toward the behavior problems of children. Beilin's review suggests greater congruence in recent years between teachers' and clinicians' attitudes. On the basis of role theory (the teachers being essentially task-oriented, the clinicians being adjustment-oriented), it is held that complete congruence is not likely to be achieved. The relationship of sex and age to attitudes toward behavior problems is also explained in the light of role theory.

Reading number 15, by Kounin and Gump—they are both professors at Wayne State University—is the report of a portion of a research project pertaining to teachers' management of children's behavior in classroom settings. In this study children's conceptions of serious misbehavior were solicited through an individual-interview technique. The conceptions of children with *punitive teachers were compared to those of children with nonpunitive teachers. The differences discovered are discussed in theoretical terms.*

*Using pupil achievement as the dependent variable and pupil affect-need, teacher warmth, and teacher permissiveness as independent variables, C. M. Christensen reports an exploratory study testing four hypotheses in reading number 16. Christensen is professor of education at the University of Alberta, Canada. The contention (inferred from data yielded by his study) that the teacher's affective response is more important for growth in achievement than permissiveness has some interesting implications for programs in teacher education.*

*Both the pattern of influence manifested by the teacher and group contagion (the tendency of a student to comply because he thinks others are complying, or not to comply if he thinks others are resisting the teacher's influence), in terms of their effects on group compliance, are investigated next by Flanders and Havumaki. At the time their research was carried out, the authors were associated with the Laboratory for Research in Social Relations at the University of Minnesota. Currently, Flanders is professor at the University of Michigan, and Havumaki is professor at Bemidji Teachers College. The authors postulate that teacher-pupil relationships are essentially superior-subordinate, the teacher being the ultimate authority. Data yielded by this experiment lead to inferences about the relative effects of teacher influence and the influences of group conformity.*

*The last selection reports a study of questions surrounding the application of principles derived from reinforcement-learning theories. The study was conducted by Loree and Koch, who were then at Louisiana State University. Loree is presently professor of educational psychology at the University of Alabama. Some interesting implications for teacher behavior in situations involving classroom interactions arise from the investigation.*

# 12

# Some Relationships Between Pupil Behavior and Certain Teacher Characteristics

## — *David G. Ryans*

*In this research report, the author concerns himself with the relationship of certain teacher to certain pupil behaviors. No attempt is made to infer causal relationships; observations are made of "behavior in process" and not of the products or outcomes of the behavior (e.g., reading achievement, emotional stability). Interesting differences are noted between the relationship of teacher behavior to pupil behavior at the elementary and secondary levels. There is a suggestion that "the way" the teacher "appears" to his class is an important variable and is related to pupil behavior at both the elementary and secondary levels.*

IT USUALLY is assumed that what goes on among pupils in a classroom is dependent to a major degree upon the presence of a teacher—a teacher whose role is to provide leadership in "learning" activities at which the pupils themselves are initiates, unprepared without assistance to determine appropriate goals or to select and employ suitable ways and means of attaining such goals. The teacher, by virtue of education and experience, is expected to be able to guide and assist his pupils in the acquisition of usable knowledge, understandings, skills, and group approved attitudes and

Reprinted from *Journal of Educational Psychology*, 1961, **52**, 82–91, by permission of the author and the American Psychological Association.

value systems. It is unthinkable, except in very rare cases, that individual children or groups of children *entirely on their own* might successfully pursue the kinds of objectives that permit adjustment to the complex social and material worlds in which we live.

Qualitatively, at least, the logic of this argument appears sound: i.e., pupil behavior is a function of teacher behavior, the teacher being a necessary, though not sufficient, condition for purposeful and productive pupil performance. Beyond this point we usually are less willing to express ourselves regarding the nature of the pupil behavior-teacher behavior relationship. We like to think specified, distinguishable teacher acts or teacher behaviors produce, or at least contribute to, certain distinguishable pupil acts or behaviors (Mitzel, 1960; Ryans, 1957, 1960). Some of us go on to postulate quantitative or functional relationships between kinds of teacher behavior and pupil behavior—assuming that more of a certain kind of teacher behavior will result in more effective pupil behavior—and proceed to gather data and evaluate the evidence (Anderson & Brewer, 1946; Christensen, 1960; Cogan, 1958; Gnagney, 1960; Jayne, 1945; Withall, 1952; and others). However, such functional relationships are not easy to demonstrate empirically, partly because of difficulties of observation and mensuration and, undoubtedly, partly because of the complex interactions among teacher and pupil characteristics (Ryans, 1956). Certainly research that might permit inferences regarding dependency relationships (producer-product relationships), with classroom situations providing the setting, is difficult to arrange and carry through—largely due to limitations imposed by practicality.

A correlational approach (less satisfactory than a dependency relations strategy for yielding the definitive kinds of answers a researcher often seeks, but nevertheless providing important information about the existence and degree of functional relationships) often poses much less of a data obtaining problem and, at the same time, provides an alternative avenue to clues and cues—in this case, regarding the influence of teacher behavior on pupil behavior.

It was this second sort of relationship (the interdependency or correlational relationship) between "what teachers do" and "what pupils of the teacher do" with which the presently reported investigation was concerned: with the degree to which certain indexes of pupil behavior tend to be functionally related to, or to manifest variance in common with, certain indexes of teacher behavior.

## Method

First of all, it is necessary to point out that the approach to pupil behavior employed in this research was a direct one, involving the immediate observation and assessment of the behavior of pupils. The pupils were observed in their regular classrooms in the presence of their teachers—teachers who *also* were observed and assessed.

Admittedly, this directly observed pupil *behavior in process* may be quite different from the products resulting from pupil behavior (Mitzel, 1960; Mitzel & Gross, 1956; Ryans, 1957, 1960). Researchers in the area sometimes indicate it is more important and relevant to judge pupil behavior in terms of its products—the measurable skills, understandings, and attitudes acquired by pupils—than merely to assess samples of ongoing pupil behavior. After all, they point out, it *is* the end product we are teaching for.

Now the method employed in the present research does involve an apparently reasonable, though certainly unproved, assumption: namely, that pupils who carry on their classroom activities in certain ways will acquire intended learnings more successfully than pupils who behave or act in other ways. But the fact remains that education is interested ultimately in pupil behavior and in teacher behavior for the results produced; immediate behavior is of interest only insofar as it is a means to an end. So, one might contend that the pupil behavior with which this paper deals is of only incidental concern. The author does not agree, but it is a point of methodology that should be made clear at the outset.

Next, it is appropriate to explain what *kinds* of directly observable pupil behavior the research dealt with.

Obviously many specific behaviors, or acts, may be noted in a classroom even during a very limited sample of time. One thing assumed in this research (a basic postulate of trait theory that makes the researcher's task tolerable) was that individual behavior possess some generality and therefore may be classified in a fairly limited number of classes or groups, each group being made up of related specific behaviors that have a common substratum which identifies the group or class. Thus, a number of pupil behaviors such as "rudeness to teacher and/or other pupils," "interruption of one another," "impatience," "refusal to participate," "quarrelsomeness," "sullenness," "disturbing noisiness," etc. might all be considered as belonging to a class of behaviors we could, for convenience, label obstructiveness. Again, pupil behaviors such as "courteousness," "friendliness," "attentiveness to advice or criticism," "diligence in completing study assignments," "initiation of constructive activity in the absence of specific direction from the teacher," etc. could be thought of as belonging to a class of behaviors we might call responsibility.

In this research we selected (after considerable study of previous classroom and "personality" research and after a good deal of trial and error involving classroom observation and assessment) and focused our attention on a limited number of "important" pupil behavior dimensions. The term "dimension" was used with intent, since in planning the research it seemed reasonable to attempt to identify and assess acts of pupils in the classroom, first, in light of their qualitative characteristics (the general class of behaviors to which a specific act belonged) and then, according to their positions on quantitative scales defined by subclasses of behaviors falling at opposite poles of the assumed continua. Thus, one bipolar dimension of pupil behavior selected because it was judged to be an important one in the classroom was the Apathetic-Alert dimension. Specific pupil behaviors that seemed to contribute to alertness (e.g., "responded eagerly," "appeared anxious to recite and participate," "watched teacher attentively," etc.) defined one end of the continuum and pupil acts that appeared to contribute to apathy ("listless," "restless," "wandering attention," "indifference," etc.) defined the other.

TABLE 1. Loadings of Observed Dimensions of Pupil Behavior and Teacher on a Pupil Behavior Factor: Secondary School Classes
(From Ryans and Wandt, 1952)

| | Oblique Factor | | | | | |
|---|---|---|---|---|---|---|
| Classroom Behavior Dimension | A | B | C | D | E | F |
| Pupil Behavior: | | | | | | |
| Apathetic-Alert | .07 | .01 | .53 | .21 | .05 | —.05 |
| Obstructive-Responsible | —.02 | .03 | .46 | .04 | —.06 | .10 |
| Uncertain-Confident | .17 | .02 | .43 | —.01 | .17 | —.03 |
| Dependent-Initiating | .00 | —.05 | .55 | .04 | .21 | —.02 |
| Teacher Behavior:[a] | | | | | | |
| Partial-Fair | | | —.04 | | | |
| Autocratic-Democratic | | | —.02 | | | |
| Aloof (G)-Responsive | | | .25 | | | |
| Aloof (I)-Responsive | | | .02 | | | |
| Restricted-Understanding | | | .08 | | | |
| Harsh-Kindly | | | —.02 | | | |
| Dull-Stimulating | | | .32 | | | |
| Stereotyped-Original | | | .17 | | | |
| Apathetic-Alert | | | .12 | | | |
| Unimpressive-Attractive | | | .01 | | | |
| Inarticulate-Articulate | | | —.04 | | | |
| Montonous (V)-Pleasant | | | —.01 | | | |
| Evading-Responsible | | | —.08 | | | |
| Erratic-Steady | | | .08 | | | |
| Excitable-Poised | | | .06 | | | |
| Uncertain-Confident | | | .20 | | | |
| Disorganized-Systematic | | | .02 | | | |
| Inflexible-Adaptable | | | .00 | | | |
| Pessimistic-Optimistic | | | .19 | | | |
| Immature-Integrated | | | .20 | | | |
| Narrow-Broad | | | —.06 | | | |

NOTE: $N = 249$ senior high school mathematics, science, English, and social studies classes.

[a] Loadings of teacher behavior dimensions are omitted here for factors on which the loadings of pupil behavior dimensions were not pronounced.

In attempting to assess pupil behavior, trained observers employed several such dimensions, assigning each class of pupils observed a value, on a scale extending from 1 to 7, on each dimension. An assessment of 1 (at the left pole) on the Apathetic-Alert dimension indicated an inference presumably based upon observation of a preponderance of apathetic-type behaviors among the pupils; a 7 (at the right pole) indicated an inference based upon the presumed observation of many alert-type pupil behaviors. A "Glossary," which accompanied the assessment form and which was employed in the training of observers, helped to standardize the assessment procedure.

TABLE 2. Loadings of Observed Dimensions of Pupil Behavior and Teacher Behavior on Two Factors Involving, in Part, "Pupil Participation" and "Controlled Pupil Activity": Elementary School Classes
(From Ryans, 1952)

| Classroom Behavior Dimension | Oblique Factor A | B | C | D | E |
|---|---|---|---|---|---|
| Pupil Behavior: | | | | | |
| Disinterested-Alert | .30 | .50 | .11 | .08 | .00 |
| Obstructive-Constructive | .16 | .63 | —.07 | .07 | .04 |
| Restrained-Participating | .40 | .10 | .04 | .00 | .18 |
| Rude–Self-Controlled | —.01 | .66 | .10 | .16 | —.05 |
| Apathetic-Initiating | .60 | .29 | .05 | —.08 | .02 |
| Dependent-Responsible | .46 | .49 | .19 | —.03 | —.10 |
| Teacher Behavior:[a] | | | | | |
| Partial-Fair | .11 | .06 | | | |
| Autocratic-Democratic | .42 | —.11 | | | |
| Aloof (G)-Responsive | .00 | —.09 | | | |
| Restricted-Understanding | .27 | —.06 | | | |
| Unattractive-Attractive | .00 | .02 | | | |
| Disorganized-Systematic | —.01 | .56 | | | |
| Inarticulate-Fluent | .00 | .33 | | | |
| Inflexible-Adaptable | .36 | .02 | | | |
| Harsh-Kindly | .25 | .08 | | | |
| Apathetic-Alert | .05 | .08 | | | |
| Aloof (I)-Responsive | —.02 | —.10 | | | |
| Stereotyped-Original | .52 | .14 | | | |
| Changeable-Constant | —.04 | .54 | | | |
| Excitable-Calm | —.00 | .36 | | | |
| Uncertain-Confident | —.02 | .44 | | | |
| Irresponsible-Responsible | .04 | .60 | | | |
| Pessimistic-Optimistic | .05 | .04 | | | |
| Infantile-Mature | .21 | .44 | | | |

NOTE: $N = 275$ third and fourth grade elementary school classes.

[a] Loadings of teacher behavior dimensions are omitted here for factors on which the loadings of pupil behavior dimensions were not pronounced.

Estimates of the reliability of assessment of the pupil behavior dimensions, based on correlations between independent assessments of two observers visiting the same classes at different times ranged from .43 to .65 for 1,518 elementary school classes observed, and from .43 to .63 for 1,911 secondary school classes.

Teacher behavior, observed and assessed following the same procedure as that described for pupil behavior, was considered in light of dimensions such as Harsh-Kindly, Disorganized-Systematic, Dull-Stimulating, etc. (see Tables 1 and 2 above). Reliability estimates of the assessments of the several dimensions of observed teacher behavior, based on correlations of the assessments of the first observer and second observer, generally were between .50 and .60—for elementary teachers coefficients for the different dimensions clustered around .55, and for secondary school teachers, around .60.

A minimum of two independent sets of assessments was obtained for each class of pupils and for each teacher, the separate ratings subsequently being weighted equally (by transformation to a scale employing a common mean and standard deviation) and combined to form composite assessments for each teacher and each class of pupils on each of the bipolar dimensions considered.

The teachers and classes of pupils upon which data of this report are based, generally, were from the school systems of fairly large cities. Geographically, the midwest and west accounted for a majority of the school systems sampled (93%). The classes represented 274 elementary schools and 103 secondary schools. Because of dependence upon the voluntary cooperation both of administrative offices responsible for school systems and of individual teachers, the sample was not selected so as to constitute a known probability sample.

## Results

Estimates of pupil behavior and of teacher behavior, obtained by the direct observation procedures described above, were subjected to correlational analyses. It should be noted that the findings, therefore, have to do simply with concomitance and permit inferences of interdependency relationships only. Inferences regarding "antecedent-consequent" or "producer-product" relationships may be suggested by such data, but the design employed precludes the testing of dependency relationship hypotheses.

### Relationships among the Pupil and Teacher Behavior Dimensions

One set of findings to be reported has to do with the extent to which the observed pupil behavior dimensions were related to the concurrently observed teacher behavior dimensions.

Here, in fairness to inferences that may be drawn, it should be noted that the pupil behaviors of the class of a given teacher were observed and assessed during the same session, or period, during which the teacher of that class was observed and assessed. Obviously there well may be contamination of one set of assessments by the other. (It would have been possible to at least partially correct for this, but it also would have added materially to both dollar and time expenditures; consequently, the choice was to recognize this source of error variance, but not to attempt to control it.)

In reviewing the pupil behavior-teacher behavior relationship findings the results for elementary teachers and secondary teachers will be presented separately; inferences that may be drawn are somewhat different. The interaction of grade level with relationships of this sort appears to be pronounced.

*Relationships in the secondary school.* In the secondary school (Ryans & Wandt, 1952), all of the pupil behavior dimensions considered in this research tended to be interrelated—the intercorrelations ranging from .42 to .68. They tended to hang together as a single cluster, all contributing to the same oblique factor when the intercorrelations of pupil behavior and teacher dimensions were factor analyzed. Table 1 shows relevant factor loadings. It may be noted from Table 1 that the factor defined by the distinctly homogenous cluster of pupil behavior dimensions was *not* significantly contributed to by the teacher behavior dimensions assessed, with one exception (the Dull-Stimulating teacher behavior dimension). Of all the teacher behavior dimensions, *only* that which had to do with the extent to which a teacher was judged Dull-Stimulating seemed to be closely associated with the pupil behavior dimension cluster.

In the secondary school, pupil behavior (as here defined and assessed) seemed not to be as closely related to teacher behavior as we sometimes assume it to be.

*Relationships in the elementary school.* In the elementary school (Ryans, 1952) we note a distinctly different situation. Table 2 presents results comparable to those shown for the secondary school in the preceding table. With the elementary school data, however, the tight clustering of pupil behavior dimensions, relatively independent of teacher behavior dimensions (noted in the high school), is absent. Instead, the pupil behavior dimensions investigated—with intercorrelations ranging from .21 to .65—split up and fell into one or the other of two loosely correlated (.08) factors, or groups, of *combined* pupil behavior and teacher behavior dimensions.

In elementary school classrooms, the loadings of the pupil behavior dimensions thus fell on two somewhat different factors, rather than on a single one, as was the case when secondary school classrooms were considered.

The pupil behavior dimensions contributing to the first factor appeared to be ones having to do with pupil participation. And related to this set of pupil dimensions were those teacher behavior dimensions that had been labeled Stereotyped-Original, Autocratic-Democratic, and Inflexible-Adaptable—traits not entirely unlike the Dull-Stimulating teacher behavior dimension in the secondary school analysis. These kinds of relationships appear to make sense—common, as well as statistical.

The pupil behavior dimensions, which as a group might be referred to as *controlled* pupil activity, were loaded on a second factor, which also included such teacher behavior dimensions as Irresponsible (evading)-Responsible, Disorganized-Systematic, Changeable-Constant, Infantile-Mature, and Uncertain-Confident—teacher characteristics that suggest disorganized, irresponsible vs. businesslike, responsible teacher behavior.

In the elementary school, then, pupil behavior and teacher behavior seemed to be more noticeably *inter*dependent than in the secondary school, with *participating* pupil behavior seeming to be related to flexible, original, democratic teacher behavior and *controlled* pupil behavior to systematic, responsible teacher behavior.

It also may be noted parenthetically that the analyses undertaken in connection with this research yielded evidence suggestive of a moderate positive relationship between the factors contributed to by the pupil behavior dimensions and those teacher behavior dimensions that loaded on a factor which seems to refer to "the way the teacher appears" to his or her class (teacher behavior dimensions such as Monotonous-Pleasant [voice], Unimpressive-Attractive, Inarticulate-Articulate, and Uncertain-Confident contributed to this teacher appearance pattern). This sort of relationship between pupil behavior and teacher appearance seemed to hold to some degree among both elementary and secondary school classes. In the secondary school, the median correlation between the teacher behavior dimensions that contributed to teacher appearance and the pupil behavior dimensions was .27; and the correlation between the teacher appearance factor and the factor dominated by pupil behavior (Factor C of Table 1) was .33. In the elementary school, the median correlation between assessments on the teacher behavior dimensions that related to teacher appearance and assessments of the pupil behavior dimensions was .37; and the correlations between the teacher appearance factor and the factors contributed to by the pupil behavior dimensions (Factors A and B of Table 2) were .57 and .23, respectively.

### Relationships Based on an Overall Index of Pupil Behavior

Now, to turn to another related aspect of the analyses which supports the foregoing findings, although coming at the problem in a slightly different manner:

In an extension of the studies just discussed, but with new independent samples of teachers and classes, an index (based on the combination of assessments for teacher behavior dimensions that loaded most highly on factors generated by the earlier factor analyses—Ryans, 1952, and Ryans & Wandt, 1952) was obtained for each teacher on each of the three prominent patterns of teacher behavior, and the teacher behavior pattern scores then considered in relation to pupil behavior indexes (a composite score for each class—$P_o$—obtained from the assessments assigned the pupil behavior dimensions)[1] of the teachers' classes.

The teacher behavior patterns considered were ones suggested by the earlier factor analyses: (*a*) aloof, restricted, egocentric vs. kindly, understanding, warm teacher behavior; (*b*) evading, unplanned, slipshod vs. responsible, systematic, businesslike teacher behavior; and (*c*) dull, routine vs. stimulating, imaginative teacher behavior.

Reliabilities of the indexes (scores) for the patterns of teacher behavior just named (based on correlations between the indexes yielded by assessments of different observers who independently assessed the same teachers) ranged from .70 to .80. Reliability estimates of the pupil behavior index were slightly lower—e.g., for one sample of 98 classes, when the pupil behavior dimensions were separately summed for two observers and the two sets of resulting $P_o$ values correlated, reliability coefficients of .62, .84, .72, and .64 were obtained, respectively, when the observations were made approximately 30 minutes apart, more than 30 minutes apart but during the same half-day, during different halves of the day (forenoon or afternoon) but the same day, and at least one month apart but during the same school year.

Correlations between pupil behavior and patterns of teacher behavior based on approximately 1,000 elementary school classes and a like number of secondary school classes are presented in Table 3.

*Secondary school relationships.* When the teacher behavior pattern scores were correlated with the pupil behavior index for secondary school classes of mathematics, science, English, and social studies, the evidence suggested low positive relationships between pupil behavior and teacher behavior, the obtained correlation coefficients between the pupil behavior index and the three teacher behavior pattern scores being between .07 and .26. It probably may be inferred that the relationship is not pronounced in secondary school classes.

---

[1] Assessments of each observer on each teacher behavior dimension were transformed into standard scores. Then for each pattern of teacher behavior ($X_o$, $Y_o$, $Z_o$) the standard scores of the several components were summed to provide the three pattern scores—based on the observer's assessments of a particular teacher. These pattern scores were, in turn, transformed into standard scores for each observer; and standard scores for independent observers were combined to yield a composite assessment on each of the behavior patterns for each teacher.

The pupil behavior index was obtained simply by (*a*) summing for each class, each observer's assessments (previously transformed to standard scores) on the several pupil behavior dimensions and (*b*) computing the mean of the equally weighted indexes contributed by the several observers who assessed a particular class of pupils.

TABLE 3. Correlations between the Pupil Behavior Index and Teacher Behavior Patterns

| Sample | $r_{P_oX_o}$ | $r_{P_oY_o}$ | $r_{P_oZ_o}$ |
|---|---|---|---|
| Elementary school classes: | | | |
| 834 Grades 1–6 classes | .82 | .80 | .75 |
| 144 Grades 1–6 classes | .83 | .78 | .80 |
| Secondary school classes: | | | |
| 497 Mathematics and science classes | .20 | .18 | .21 |
| 568 English and social studies classes | .18 | .21 | .26 |
| 114 Mathematics, science, English, and social studies classes | .07 | .11 | .14 |

NOTE: $P_o$ = pupil behavior (composite).
$X_o$ = kindly, understanding vs. aloof, restricted teacher behavior.
$Y_o$ = responsible, businesslike vs. evading, unplanned teacher behavior.
$Z_o$ = stimulating, imaginative vs. dull, routine teacher behavior.

*Elementary school relationships.* With elementary school classes, the pupil behavior index was found to be uniformly highly correlated with all three teacher behavior pattern scores (again, as suggested by the analyses reported in the preceding sections of this paper), the correlation coefficients ranging from .75 to .83. (It may be observed that these pupil behavior-teacher behavior correlations in the elementary school were as high as the reliability coefficients for the several pattern scores.)

Thus, for the kinds of teacher behavior and pupil behavior described, the two separate analyses support the same conclusion: i.e., that teacher behavior and pupil behavior show substantially more *inter*dependence in the elementary school as compared with the secondary. There also is the suggestion that of the teacher behavior dimensions and patterns studied, Dull-Stimulating teacher behavior may be more closely associated with secondary school pupil behavior.

Such findings, it should be recalled, are based upon correlational analysis, permitting interdependency, but not dependency (producer-product), inferences about the relationships involved. Furthermore, the relationships reported refer to group data—to groups of teachers and their classes—and the probability risks involved in applying conclusions to specific situations should be kept in mind.

## Summary

Relationships between trained observers' assessments of (*a*) classes of pupils and (*b*) the teachers of those classes, relative to selected dimensions of pupil and teacher classroom behavior, were studied.

For elementary school classes, *high* positive relationships were noted between observers' assessments of "productive pupil behavior" (e.g., assessments presumed to reflect pupil alertness, participation, confidence, responsibility, and self-control, initiating behavior, etc.) and observers' assessments of previously identified patterns of teacher behavior which seemed to refer to understanding, friendly classroom behavior; organized, businesslike classroom behavior; and stimulating, original classroom behavior.

For secondary school classes, *low* positive relationships appeared to obtain between productive pupil behavior and the above named categories of teacher behavior, with a tendency for the stimulating, original teacher classroom behavior pattern to show a slightly higher correlation with pupil behavior than the understanding, friendly or the organized, businesslike teacher behavior patterns.

The approach was correlational, permitting inferences of interdependency, but not producer-product, relationships.

## REFERENCES

Anderson, H. H., & Brewer, J. E. Studies of teachers' classroom personalities: II. Effects of teachers' dominative and integrative contacts on children's classroom behavior. *Appl. psychol. Monogr.*, 1946, No. 8.

Christensen, C. M. Relationships between pupil achievement, affect-need, teacher warmth, and teacher permissiveness. *J. educ. Psychol.*, 1960, **51**, 169–174.

Cogan, L. Behaviors of teachers and the productive behavior of their pupils. *J. exp. Educ.*, 1958, **27**, 89–124.

Gnagney, W. J. Effects on classmates of a deviant student's power and response to a teacher-exerted control technique. *J. educ. Psychol.*, 1960, **51**, 1–8.

Jayne, D. D. A study of the relationship between teaching procedures and educational outcomes. *J. exp. Educ.*, 1945, **14**, 101–134.

Mitzel, H. E. Criteria of teacher effectiveness. In *Encyclopedia of educational research*. New York: Macmillan, 1960. Pp. 1481–1486.

Mitzel, H. E., & Gross, Cecily F. A critical review of the development of pupil growth criteria in studies of teacher effectiveness. (Research Series 31.) New York: Board of Higher Education, Office of Research and Evaluation, 1956.

Ryans, D. G. A study of criterion data: A factor analysis of teacher behaviors in the elementary school. *Educ. psychol. Measmt*, 1952, **12**, 333–344.

Ryans, D. G. Theory development and the study of teacher behavior. *J. educ. Psychol.*, 1956, **47**, 462–475.

Ryans, D. G. Notes on the criterion problem in research, with special reference to the study of teacher characteristics. *J. genet. Psychol.*, 1957, **91**, 33–61.

Ryans, D. G. Teacher characteristics in relation to marital status of the teacher. *Calif. J. educ. Res.*, 1959, **10**, 68–72.

Ryans, D. G. Prediction of teacher effectiveness. In *Encyclopedia of educational research*. New York: Macmillan, 1960. Pp. 1486–1491.

Ryans, D. G., & Wandt, E. A factor analysis of observed teacher behavior in the secondary school: A study of criterion data. *Educ. psychol. Measmt,* 1952, **12**, 574–586.

Withall, J. Assessment of the social-emotional climates experienced by a group of seventh graders as they moved from class to class. *Educ. psychol. Measmt,* 1952, **12**, 440–451.

# 13

# A Study of the Effects of Teacher Knowledge of Pupil Characteristics on Pupil Achievement and Attitudes Towards Classwork

— *Kenneth B. Hoyt*

*The author describes the results of an investigation on the effect of teachers' knowledge of students' characteristics on student achievement and attitudes. The investigation was conducted over a period of six months during which time it was determined that providing information about students to teachers did, in fact, increase the teachers' knowledge of those students. However, this knowledge was not found to act preferentially on either sex or ability levels in terms of its effect on increases in student achievement in the three subject areas used—mathematics, social studies, and English. However, there was a definite tendency for increases in teacher knowledge of student characteristics to improve student attitudes toward teachers.*

TO SAY that individual differences exist in school children has little significance simply as an observation. The significance lies in the actions one can take as a result of having made this observation and the results which ensue in terms of changes in students.

Reprinted from *Journal of Educational Psychology*, 1955, **46**, 302–310, by permission of the author and the American Psychological Association.

Many people prominent in education have, over the last fifty years, urged teachers to discover the ways in which and the degrees to which school children differ in various characteristics (*1*, *2*). Some have been so bold as to state specifically such beliefs as "There is no doubt that better teaching results from supplying teachers with pertinent facts about all their pupils." (*6*).

## The Problem

One approach to the problem of relative advisability or inadvisability of providing teachers with information concerning pupil characteristics is to study some of the effects resulting from such knowledge on the part of teachers. This investigation was concerned with studying the effects of teacher knowledge of pupil characteristics on pupil achievement and teacher-pupil relationships. Seven major questions were devised which served as the focus for the investigation. These were as follows:

1. What is the effect of providing teachers with systematic appraisal data relative to pupil characteristics on teacher knowledge of pupil characteristics?
2. What is the effect of teacher knowledge of pupil characteristics on pupil achievement?
3. Does the effect of teacher knowledge on pupil achievement act differentially for boys and girls?
4. Does the effect of teacher knowledge on pupil achievement act differentially on students from different learning ability levels?
5. What is the effect of teacher knowledge of pupil characteristics on pupil attitudes towards teachers?
6. Does the effect of teacher knowledge on pupil attitudes towards teachers act differentially for boys and girls?
7. Does the effect of teacher knowledge on pupil attitudes towards teachers act differentially on students from different learning ability levels?

The major experimental factors studied with reference to the treatment effect were teacher knowledge, sex, and ability level. The major experimental variables were pupil achievement and pupil attitudes towards teachers.

## Populations Studied and Sampling Methods Employed

The experiment was conducted during the school year 1953–54, utilizing eighth-grade students in English, mathematics, and social studies classes from two public junior high schools in Minnesota. In School A, two hundred and ninety students were included on the roster and in School B, three hundred and fifty-six students. The actual experimental period was six months in length and all findings should be interpreted with this limitation in mind.

Individual students in each school were selected for and assigned to experimental sections by the method of random sampling within sex and scholastic aptitude sub-populations. Six sub-populations were formed from the roster in each school with boys and girls each being divided into highest, middle, and lowest ability groupings on the basis of obtained IQ scores on the *Otis Quick Scoring Mental Ability Test,* Beta level, Form Cm., which had been administered to all students in both schools during the school year 1952–53.

In School A, ninety students were selected and assigned at random, thirty to each of three experimental sections. Complete analysis data were available for eighty-five of these ninety students at the conclusion of the experimental period. In School B, one hundred and twenty students were selected and assigned at random, forty to each of three experimental sections, with complete analysis data being available for one hundred and nine students. The effects of the loss of these students as well as a graphical representation of the sub-populations involved can be obtained from an examination of Tables 1 and 2.

TABLE 1. Numbers of Students, by Sections, in Each Stratum for Which Complete Analysis Data Were Available in School A

| | Section 8A | | Section 8B | | Section 8C | | Totals | |
|---|---|---|---|---|---|---|---|---|
| | Girls | Boys | Girls | Boys | Girls | Boys | Girls | Boys |
| High IQ | 5 | 5 | 5 | 4 | 5 | 5 | 15 | 14 |
| Middle IQ | 4 | 5 | 5 | 4 | 6 | 4 | 15 | 13 |
| Low IQ | 5 | 5 | 5 | 4 | 5 | 4 | 15 | 13 |
| Totals | 14 | 15 | 15 | 12 | 16 | 13 | 45 | 40 |

TABLE 2. Numbers of Students, by Sections, in Each Stratum for Which Complete Analysis Data Were Available in School B

| | Section 8D | | Section 8E | | Section 8F | | Totals | |
|---|---|---|---|---|---|---|---|---|
| | Girls | Boys | Girls | Boys | Girls | Boys | Girls | Boys |
| High IQ | 7 | 5 | 7 | 5 | 7 | 6 | 21 | 16 |
| Middle IQ | 5 | 8 | 6 | 6 | 7 | 6 | 18 | 20 |
| Low IQ | 6 | 6 | 6 | 6 | 5 | 5 | 17 | 17 |
| Totals | 18 | 19 | 19 | 17 | 19 | 17 | 56 | 53 |

## Specification and Assignment of the Treatment Variables

Three distinct treatments were applied to each section during the experimental period. These were arranged for both schools in such a way that any one teacher was responsible for a different treatment with each section.

Definitions of the three treatment variables are presented in terms of the amount of information concerning pupil characteristics given to teachers.

The treatment of 'No Information' (N) is defined as the process of giving the teacher only the names of the pupils and no other systematic information.

The treatment of 'Test Scores' (T) is defined as the process of giving the teacher the pupils' test scores. These consisted of converted IQ scores from the *Otis Quick Scoring Mental Ability Test,* Beta level, Form Cm. (which all students took during the school year 1952–53); scaled scores, by parts, for the *Coöperative English Test,* Lower Level, Form Z; the *Coöperative Social Studies Test for Grades 7, 8, and 9,* Form X; and the *Coöperative Mathematics Test for Grades 7, 8, and 9,* Form X (which each student took in September, 1953). No other systematic information was given the teacher.

The treatment of 'Tests Plus Other Information' (TO) is defined as the process of giving the teacher, for each student, all the information described for the Test Scores treatment plus a pupil information blank which was administered in September, 1953, and constructed by the experimenter. In addition, each teacher was encouraged to discover all he possibly could about the characteristics of each student receiving this treatment.

These treatments were assigned to experimental sections by subject using a table of random numbers compiled by Fisher and Yates (3). The resulting assignments for School A are shown in Table 3 and for School B in Table 4.

TABLE 3. Assignment of Treatment Variables to Experimental Sections in School A

| Section | Teacher 1 (English) | Teacher 2 (Mathematics) | Teacher 3 (Social Studies) |
|---|---|---|---|
| 8A | T | TO | N |
| 8B | TO | N | T |
| 8C | N | T | TO |

TABLE 4. Assignment of Treatment Variables to Experimental Sections in School B

| Section | Teacher 4 (English) | Teacher 5 (Mathematics) | Teacher 6 (Social Studies) |
|---|---|---|---|
| 8D | TO | TO | N |
| 8E | N | N | T |
| 8F | T | T | TO |

## Description of Tests and Forms Used

Student achievement was measured in September and again in April using the *Coöperative English Test,* Lower Level, Form Z, the *Coöperative Mathematics Test for Grades 7, 8, and 9,* Form X, and the *Coöperative So-*

*cial Studies Test for Grades 7, 8, and 9*, Form X. While these tests measure recall and recognition, it is apparent that what some would regard as the relatively more important goals of retention and transferability of learned materials to new situations is not measured by these instruments. This serves to specify one limitation of the study.

Information relative to other pupil characteristics which were considered as important aspects of treatment TO were collected in September by administration of the pupil information blank.

Student attitudes towards particular teachers were measured by use of an inventory constructed by the experimenter. It consisted of items taken from the *Student Reaction Inventory* (4), contained in category 7, "Individualization of Instruction," which, on an item analysis, showed differences between 'good' and 'poor' teachers (5) significant at the 1% level. Forty such items were selected and five more non-discriminating items which had high face validity were added making forty-five items in all. Responses typical of those students make about 'good' teachers were assigned a weight +1; those associated with 'poor' teachers a weight of −1; and all items answered 'undecided' were scored as zero. This instrument was administered to each student three times, once in each of the three experimental groups to which he was assigned.

Each teacher was asked to complete a blank consisting of questions designed to measure the degree to which he was able to recall specific characteristics of certain randomly selected pupils in the various experimental sections. Six students, one from each stratum, were picked at random from each experimental section in a given school and the three teachers involved were asked to complete the blanks for these eighteen students.

## Experimental Procedures

The experimental procedures employed in this investigation can perhaps be best described in terms of controls applied during the experimental period to the various treatment procedures. Rigid controls were also applied to the collection of data, but these are not described here.

*To maintain the limitations prescribed by treatment N* teachers were urged, for each member of the section where this treatment was applied, to: (a) Refrain from examination of the pupil's cumulative record folder; (b) Refrain from discussion of this pupil in conversations with other school personnel; (c) Refrain from discussing with the student any topics which might lead to information concerning his activities outside of the classroom; and (d) Refrain from seeking out parents of the student for consultation and discussion.

*To carry through the procedures prescribed by treatment T* teachers were urged, for each member of the section where this treatment was applied, to maintain the same limitations described for treatment N. They were given copies of all initial achievement test scores and IQ scores for individual students in this section. In addition, graphs of distributions of these scores were supplied as was a scattergram showing the relationships between intelligence and achievement test scores in the subject the teacher was teaching.

*To carry through the procedures prescribed by treatment TO* the teachers were supplied, for each member of the section where this treatment was applied, with all of the information described for the T treatment. In addition, they were given: (a) Individual record folders for each student in which they were encouraged to place records of all information learned regarding characteristics of the student; (b) The pupil information blanks which the student had completed and on which the experimenter had marked some interpretive suggestions; and (c) A four-page group analysis of the section made on the basis of information obtained from the pupil information blanks.

*To work towards a clear differentiation of treatments* individual conferences were held with teachers at least once every other week during the experimental period. Here, in addition to providing frequent reminders of the restrictions imposed by Treatments N and T, the experimenter worked with each teacher in interpreting the significance of the data compiled for students in the T and TO sections. No attempt was made to force the teacher to discuss every pupil; rather, the choice of pupils to be discussed was always left with the teacher.

*To work towards the prevention of gross changes in teaching method* teachers were asked to: (a) Use the same textual material for all experimental sections; (b) Cover the same units at approximately the same time with all experimental sections; (c) Use the same general approach in teaching method with all experimental sections, i.e., if a topic were presented by means of a student panel discussion in one section, this same approach should be used in all sections. Checks made during and at the conclusion of the experimental period satisfied the investigator that these restrictions were being carried out.

## Statistical Design and Methods of Analysis

The statistical design utilized for analysis of achievement test scores is defined as $2 \times 2 \times 3 \times 3$ factorial. Specific factors involved included two schools, two sexes, three ability levels, and three treatments. This design was applied three times, once for each of the three school subjects—English, social studies, and mathematics. The method of statistical analysis employed was the analysis of variance and covariance with unequal frequencies in the subclasses using unweighted means. The method of unweighted means rather than the least squares method was chosen primarily because of the interest of the investigator in examining first-order interaction effects of schools, sex, and ability with the Treatment effects in the experimental design. Where analysis of covariance was employed, adjustment was made on the basis of initial achievement test scores.

The statistical design utilized for analysis of attitude test scores is defined as a 2 × 3 × 3 × 3 factorial. This design was applied twice, once for School A and once for School B. Specific factors included two sexes, three school subjects, three ability levels, and three treatments. The method of statistical analysis employed was the analysis of variance with unequal frequencies in the subclasses using unweighted means.

The statistical design utilized for analysis of teacher knowledge scores is defined as a 3 × 3 factorial with three teacher subjects and three treatments representing the factors involved. This design was applied twice, once for School A and once for School B. The method of statistical analysis employed was the analysis of variance with equal frequencies in the subclasses.

The assumptions involved in the use of these various statistical tools were tested separately for each design and found to be satisfied.

## Summary of Findings

1. Teacher knowledge of pupil characteristics was increased through the use of appraisal data. Significant F ratios were obtained on this analysis for mean treatment differences, for mean teacher-subject differences, and for mean interaction differences of treatment with the various teacher-subject combinations. The mean square for the treatment effect provided a significant F ratio when compared with the mean squares for error, teacher-subject, and interaction thus establishing itself as probably constituting a true effect. This finding indicates that the Treatment procedure did operate as intended.

2. Teacher knowledge of pupil characteristics was not found by itself to result in increases in pupil achievement in mathematics, social studies, or English. The application of the various treatment effects did not produce any significant differences in adjusted mean achievement test scores in mathematics and social studies. While significant adjusted mean differences in favor of Treatment N were found for adjusted achievement test scores in English, the complexities of the interaction effects present prevent one from regarding the observed differences as true treatment effects.

3. Teacher knowledge of pupil characteristics was not found to act preferentially on either sex in terms of its effect on increases in pupil achievement

in mathematics, social studies, or English. There were no significant F ratios obtained for the interactions of the sex with the treatment factor.

4. Teacher knowledge of pupil characteristics was not found to act preferentially on ability levels of pupils in terms of its effect on increases in pupil achievement in mathematics, social studies, or English. The interaction effects of ability level with the treatment factor were not significant in mathematics and social studies classes. A significant F ratio was found for the interaction between ability level and treatment in English classes. However, because of significant interactions also present between treatment and schools and between ability and schools plus significant main effects for school and treatments, the exact interpretation of the significant interaction of ability with treatment was not clear.

5. There was a definite tendency for increases in teacher knowledge of pupil characteristics to improve pupil attitudes towards teachers. Significant differences in favor of Treatment TO were found in School B. In School A, the F ratio of 1.461 was smaller than that of 3.04 required to reject the hypothesis of equality at the 5% level of significance. However, the mean of 8.37 for Treatment N, 8.77 for Treatment T, and 12.12 for Treatment TO were all in the direction indicated by the significant findings in favor of Treatment TO which were found in School B.

6. Teacher knowledge of pupil characteristics was not found to act preferentially on either sex in terms of its effect on pupil attitudes toward teachers. The interaction of sex with treatment was not significant in either school.

7. Teacher knowledge of pupil characteristics was not found to act preferentially on ability levels of pupils in terms of its effect on pupil attitudes toward teachers. The interaction of ability level with treatment was not significant in either school.

## REFERENCES

1. Warren, Baller, *et al.* Essential public personnel records. *Teach. coll. Rec.*, January 1947, 268–286.
2. Bush, R. N. A. Study of student-teacher relationships. *J. educ. Res.*, May 1942, 645–656.
3. Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural and medical research.* New York: Hafner, 1949.
4. Grim, P. R., & Hoyt, C. J. Appraisal of teaching competency. *Educ. res. Bull.*, April 16, 1952, 85–91.
5. Grim, P. R., & Hoyt, C. J. Excerpts from two instruments for appraising teaching competency. *J. educ. Res.*, May 1953, 705–710.
6. Wadsworth, R. D. Give teachers the facts about pupils. *Calif. J. secondary Educ.*, February 1946, 84–87.

# 14

# Teachers' and Clinicians' Attitudes Toward the Behavior Problems of Children: A Reappraisal

## — *Harry Beilin*

*The present author reviews a group of studies that compare teachers' and clinicians' (i.e., psychiatrists, psychologists, and social workers) rankings of the importance of selected behaviors as indications of present or potential maladjustment. Beilin compares the results of recent studies with those of Wickman's landmark study, published in 1928. At that time, Wickman reported a negative correlation between the rankings of behavior problems by mental hygienists and by teachers. Although subsequent research has continued to demonstrate differences in teachers' and clinicians' attitudes toward these problems, the differences have grown smaller. Beilin accounts for this change in terms of role theory. Suggested reasons for the greater congruence today between teachers' and clinicians' attitudes include differences in teacher training and increased emphasis on such things as social-adjustment education in the classroom. These have led to increased attention given by teachers to motivational factors in behavior and to the social pressures acting on children.*

THE CONTRAST between contemporary American education and that of 40 or 50 years ago is striking in at least one respect, the influence of

Reprinted from *Child Development,* 1959, **30,** 9–25, by permission of the author and the Society for Research in Child Development.

psychology, in particular, clinical psychology.[1] It would require little effort to detail the many and diverse ways teacher training, parent education, and child care reflect the consequences of psychology's influence. The future historian will undoubtedly dwell upon the part played by E. K. Wickman's 1928 Commonwealth Fund monograph, "Children's Behavior and Teachers' Attitudes," in this development (51). Wickman's report, which contrasts teachers' and "mental hygienists'" attitudes toward the behavior problems of children evoked an assault upon the teacher's mode of dealing with children when it made evident that teachers' attitudes were widely at variance with those of clinicians. The effect of its publication is still felt. The contiguity of events might suggest to some a causal relationship between widespread knowledge of the Wickman findings and the emergence of clinical psychology as a force in contemporary education. However, a more temperate and realistic appraisal would accept the ubiquitous penetration of psychology into American life and not as a condition unique to education. The impact of the monograph was in actuality only one of a series of challenges to the values and attitudes of educational personnel. But irrespective of its true role—whether as reflection, or initiator of a Zeitgeist—the place of the Wickman study in education and child psychology has been significant and will probably continue to be.

Wickman's results, in the main, suggested that mental hygienists were primarily concerned with *withdrawing* and other nonsocial forms of behavior in children of elementary school age, whereas teachers of these same children were more concerned with *classroom management, authority,* and *sex problems.* The results influenced many (starting with Wickman) to urge teachers to adopt a hierarchy of attitudes closer to that of the clinician. This view presumed that the clinician's judgment should be accepted as the criterion for adequate and inadequate behavior. Few have challenged this thesis.

The intent of the present review is to examine what the result of 30 years of research suggests for continued acceptance of this point of view. To anticipate, it will be suggested that Wickman's findings be reinterpreted and his prescription for change in educational policy modified.

## The Wickman Study

The Wickman studies were begun in Minneapolis in 1924, but a more ambitious program was undertaken in Cleveland in 1925–1926, where the following was done:

---

[1] Although the period following the first World War saw the impact of the testing movement and the effects of Behaviorism, it was not till the 1930's and 1940's that clinical psychology became a part of the child development and educational scene in a major way.

1. In a single pilot school, teachers' characterizations of undesirable behavior, with indications of "sensitiveness" to their occurrence, were elicited by questionnaire.

2. Teachers' attitudes toward various types of problems were obtained by three measures (detailed in part below).

3. On rating scales, teachers noted their reactions to problems themselves, then to pupils in whom the problems were observed, and finally to the total adjustment of their pupils.

4. Subsequently, the teachers from 13 schools in six communities and two additional teacher groups enrolled in graduate school were studied using the rating scale method developed for the pilot study. The most important feature of this involved the rating of a number of behaviors obtained from the teachers' original freely given characterizations of problem behavior. The results are reported as mean ratings and rankings of mean ratings.

5. Finally, 30 "mental hygienists" (8 psychiatrists, 4 psychologists, 13 psychiatric social workers, and 5 teachers with social work background) from child guidance clinics in three cities were studied for their attitudes toward 50 of the same behaviors rated by teachers. The mean ratings and rankings of ratings were then contrasted and correlated with those of teachers.

The rating instructions for teachers stressed: (a) *present* problems, (b) "seriousness" of the problems or "difficulties" created by them, and (c) rapid responses to the rating scale. With clinicians, the emphasis was on (a) relevance of the problem behavior for *future* adjustment, (b) though "seriousness" and "difficulty" were retained, the focus was on the "importance" of the behavior, and (c) no time limit was imposed for response to the rating scale.

The principal results can be summarized as follows:

1. Teachers were most aware of overt and aggressive behaviors, inattention to school tasks, and behaviors which violated their standards of morality. They were much less concerned with behaviors indicative of social or emotional maladjustment not directly related to school routine.

2. Boys were reported more frequently than girls for behavior problems.

3. Teachers preferred the less active, more compliant behavior of girls to the more aggressive behavior of boys. Desirable conduct for teachers, then, took on the distinguishing characteristics of girl behavior.

4. "Mental hygienists" considered withdrawing and other nonsocial forms of behavior most serious and discounted the teachers' stress on antisocial behavior and violations of school rules.

5. There was a rank order correlation of $-.22$ between the rankings by mental hygienists of 50 behavior problems and the original Cleveland

teachers' ($N = 28$) rankings of the same behaviors. The correlation was $-.11$ when the full sample was used ($N = 511$).

These findings were interpreted by Wickman in Thorndikian stimulus-response terms. The teachers distinguish, he said, between the attacking and withdrawing types of behavior problems. Their attitudes are principally determined, however, by the attacking nature of the child's conduct. The aggressive behaviors are identified and considered more serious because the teacher is aroused to counterattack by virtue of the frustration in him. On the other hand, the responses to withdrawing forms of behavior are modified by sympathy and protective feelings.

On the basis of these findings Wickman then proposed that:

1. Teachers' attitudes should be influenced to become more like the "ideal" clinicians. (Clinicians' attitudes are considered ideal because their judgments are (ostensibly) based upon knowledge of research in child adjustment.)

2. Teacher attitudes should be changed not by exhortation but by (a) information about child behavior through seminars and other learning experiences; and (b) practice in therapy with children.

3. Teachers' functions should be less concerned with intellectual learnings and more with life adjustment.

After the appearance of the 1928 monograph some serious limitations in method and conception were pointed out by G. Watson (50) in a critical note,[2] which are as cogent now as when first offered. The majority of efforts to rectify the deficiencies have concerned only some of the criticisms. The others, however, may be of as great issue as those treated.

Watson's objections were:

1. *The procedures themselves are open to criticism.*

   a. The directions given to teachers and clinicians were not the same. Teachers were instructed to rank behaviors for *present* seriousness; clinicians, for *future* adjustment.

   b. The time given to respond to the questionnaires was not identical. Teachers were under the control of the experimenter; clinicians were allowed an extended period to respond.

   c. No definitions were given for the behaviorial terms to be rated leaving to each subject the interpretation of the terms, and thus further reducing comparability of the results.

The issues raised by Watson's concern with methodology are intimately related to other criticisms.

---

[2] One writer has even wondered how the study could have been so widely and uncritically accepted with these limitations. The answer probably rests in the fact that its thesis was part of a powerfully developing movement.

2. *The choice of mental hygienists' attitudes toward the behavior problems of children as criterion for evaluating teachers' attitudes toward the same problems is open to question.*

The Wickman study and others that follow (though not all) accept the clinicians' judgments as a criterion either implicitly or explicitly. Watson observes that there is no reason to suppose clinicians to be "correct" and teachers not, rather than vice versa. Wickman is questioned for not even considering this possibility.

3. *There has been too ready an acceptance of a causal relationship between withdrawing behavior in childhood and maladjustment in adulthood.*

In addition to questioning whether the term "withdrawing" means the same thing to teachers and clinicians, Watson questioned whether withdrawing behavior in childhood is causally related to, or predictive of, maladjustment in adulthood. Although this is contended in more than one theoretical position, there was very little evidence for the validity of this claim in 1928, and little more is available now. In Watson's paper there is reference to a pilot study which, for all its limitations, casts some doubt on the aforementioned assumption.

## The Literature

After the Watson critique there was concern for the validity of the results and replications were undertaken with one or another modification in design, ultimately making comparability difficult.[3]

The studies, in the main, fall into the following groupings according to procedures used:

A. *Studies employing descriptions of problem behavior.*
  1. Teacher nomination of children with problems, followed by description and classification of problem behaviors (*3, 5, 18, 29, 48*).
     a. In addition to all or part of the above, some use is made of a rating scale of problem behaviors (*8, 15, 24, 25, 37, 47, 53, 54*).
  2. Teacher description of problem behavior (with no reference to specific children) from which a rating scale is developed or the descriptions themselves are used (*10, 11, 22, 36, 41, 45, 51*).
  3. Children identified and described as problems by a social or therapeutic agency (*6, 33*).

B. *Studies employing the Wickman rating scales.*
  1. With Wickman's directions (*1, 20, 35*).
  2. With modifications of Wickman's directions (*14, 19, 26, 38, 39, 40*).

---

[3] We shall consider all studies found that bear upon the problems focused on by Wickman and his critics in spite of the lack of comparability. Some studies are included which antedate Wickman because the data they offer are relevant.

**Confirmation of Wickman Findings**

Early studies that made use of the Wickman scales "confirmed" what Wickman had found in that the rankings made by the teachers in other communities approximated those of Wickman's teachers. Boynton and McGaw (*8*) obtained a correlation of .87 between Wickman's and their own teacher ratings. Dickson (*13*), Laycock (*22*), Young-Masten (*53*), and Yourman (*54*) gave similar results as did the Epstein (*15*) and Snyder (*37*) studies. Thompson (*45*) offered the added information that teachers were more nearly in agreement with the rankings made by parents than with the rankings of child psychologists. In Young-Masten's study (*53*), in addition to securing conduct reports for nominated problem children, she observed a group of 28 problem children and a control group of the same number. From the classification of behavior records for each child observed, she obtained a statistically significant difference between the groups for behavior "which from the standpoint of the teacher, was annoying, and upset the order and peace of the classroom, and interfered with the other children" (p. 180).

Five years after the appearance of the Wickman report Bain (*1*) assessed the attitudes of teachers enrolled in graduate work using the original directions for teachers. The resulting rankings showed higher correlations with Wickman's mental hygienists than had been true for the original teacher sample. The correlations reflected greater teacher concern with *recessive, withdrawing behaviors* than with *active offenses* and *sex problems*.[4]

Replication of the Wickman study, employing the same directions, was undertaken in 1951 by Schruppe and Gjerde (*35*). They concluded that teachers' attitudes in 1951 agreed more closely with the "ideal criterion" (clinicians' attitudes) than did 1927 teachers' attitudes (a correlation of .57 (1951) compared with −.04 (1927) using means of teachers' and clinicians' ratings). None of the traits listed as most serious by one group was listed as least serious by the other. Wickman had five such differences. Griffiths' study (*16*), although not a replication of Wickman's, was similar to some features of it. He found that behavior difficulties most reported by teachers relate to violations of classroom rules and work rules. However, he indicated that teachers have changed over the years as evidenced by their awareness that a child who is easy to manage is not necessarily well adjusted.

Del Solar's (*11*) study is a report of interviews with parents and teachers which details the joys and problems of child rearing. It was the shy and withdrawing, rather than aggressive, behaviors which concerned this group of teachers and parents. These findings contradict Wickman and most other investigators. Del Solar's findings may result, however, from the nature of the questions asked in the interview (only one, concerning liabilities of students, is cited), the small size of the sample (6 teachers rating 28 children), and the atypical nature of the subjects (all teachers had advanced study and were employed in a school of above average socioeconomic status).

---

[4] In another part of the study Bain reports that one semester of instruction had little effect in changing attitudes toward problem behaviors. Whether the differences reported in the Bain study result from the select character of the sample is not known.

We would conclude from these studies (holding the question of the validity of Wickman's methodology aside for the moment) that there is considerable evidence to indicate agreement with Wickman's original findings. Furthermore, there has been an observable shift in the intervening years in the attitudes of teachers in the direction of being more like those of "mental hygienists." In spite of greater congruence, however, a sizeable difference remains between the attitudes of teachers and clinicians toward behavior problems of children.

### The Methodical Issue

Modifications as a rule have aimed at introducing uniformity in the administration of the problem rating scales. This has meant either modifying the directions given clinicians or teachers so that both groups could respond under the same stimulus conditions.

In Ellis and Miller's 1936 investigation (*14*) the instructions used were those originally given clinicians. In this study they were administered to teachers. With this change in method, ratings of teachers correlate .49 with Wickman's mental hygienists (and .65 with Wickman's teachers). Again, the change is the result of increased realization of the seriousness of *withdrawing* and *recessive* personality traits. The investigators note, however, that teachers "still consider" *violations of general standards of morality* and *transgressions against authority* as the most seriously rated types of behavior.

Sparks' investigation (*38*) made use of both the Wickman teacher and clinician directions and scales. These were administered to teachers and graduate students (in education). Teachers' ratings made in terms of seriousness for future adjustment (the mental hygienist form) were different from the kinds of ratings made when the directions stressed "troublesomeness" in the classroom, although in both cases they were still different from psychologists' ratings. The correlation of ratings (for teachers) between the original teacher form and the mental hygienist form of the scale was .05.

Mitchell's study (*26*) was conducted in the same cities where Wickman had done his 13 years earlier. The scales and directions were both modified, the scales but slightly. The directions given teachers and mental hygienists were the same: they were asked to rate behavior traits keeping in mind the behaviors of fifth and sixth grade children they had observed. (No such grade specification was made by Wickman.) The directions given the 1940 group were "almost" identical with those given by Wickman and these we presume were the instructions originally given clinicians. The correlation between means of ratings of 1927 and 1940 mental hygienists was .80. Some mental hygienists had apparently become more "conservative" in their ratings. They no longer considered *unsocial, withdrawing,* and other traits as extremely grave—in fact, no traits were now so considered. Whereas, the correlation in 1927 between the means of teachers and mental hygienists' ratings was .08 (by the rank difference method), the correlation in 1940 was .70. Mitchell interprets this to mean that either the identity in directions accounts for these results or teachers and clinicians have moved closer in their judgments.

In Stouffer's 1952 investigation (39) teachers were first given the original instrument for teachers; later, the original scale for mental hygienists. A group of mental hygienists were then administered the original mental hygienists' scale. Stouffer reports a correlation of .52 between the teachers' ranking of problem behaviors and mental hygienists' ranking of the same, employing different directions (the original Wickman procedure). A correlation of .61 was obtained when the instructions (those of mental hygienists) were the same for both groups. A rank order correlation of .87 was obtained between the rankings of Stouffer's and Wickman's mental hygienists.

Stouffer concluded that while teachers' attitudes toward the behavior problems of children have changed there has been little change in the attitudes of mental hygienists.[5] The changes in teachers' attitudes reflect again the reduced importance of problems related to *honesty, sex, truancy,* and *classroom order,* and increased importance of *withdrawing* and *recessive* personality traits.

Stouffer's 1956 study (40) is of particular interest because it recognizes that a difference may exist between teachers of different grade levels. In contrast to his 1952 study the later one deals with secondary school teachers. The instructions were those for clinicians. The results were contrasted with those of elementary teachers in the prior study and with Wickman's (elementary) teachers. The rank order correlation of secondary teachers' rankings and elementary teachers' rankings was .88; between the same secondary teachers' and mental hygienists' ratings, .49. As reported in the earlier study, the correlation between elementary teachers' and mental hygienists' rankings was .61. Elementary teachers are, then, in greater agreement with mental hygienists than secondary school teachers. In terms of children's behaviors, elementary teachers are more concerned with withdrawing tendencies; secondary school teachers, with classroom management and problems related to class work and school routines.

The findings that the criteria of adjustment and maladjustment differ depending upon age and grade level is given support in a study of Beilin (3). In this instance, the procedure involves teacher nomination of maladjusted children, descriptions of their distinguishing characteristics, and content analysis of the descriptions. In general, an age trend (from elementary grades to young adulthood) was found with a concern (in elementary grades) for social-interpersonal aspects of adjustment (e.g., *withdrawal, aggressiveness, emotional instability*) to later concern (in high school) with character traits (e.g., *reliability, dependability*) and finally (young adulthood) with *achievement* and *integration into the community.*

---

[5] Note that Mitchell *(26)* suggests such a change although the correlation for the same relationship differs from Stouffer's by only .07.

In the 1957 study by Hunter (19), elementary and secondary school teachers were sent the Wickman mental hygienists' rating scale and instructions. His results are similar in direction to Mitchell's and Stouffer's, though not in size of correlations. The correlation between teachers' and mental hygienists' rankings of mean ratings was .22. *Aggressiveness* is still rated more highly by teachers than by mental hygienists.

Using a procedure which raises doubts, Ullmann (47) had teachers nominate well and poorly adjusted children who were then rated on a 144-item rating scale. Discrimination indices were computed for each item. The check list was then submitted to a group of mental hygienists who rated the extent to which each item was indicative of good or poor adjustment. The discrimination indices were correlated with the means of clinicians' ratings of the items. The correlation was .86. For favorable items it was .69; unfavorable, .50. These results are interpreted by Ullmann as confirming the Mitchell findings that teachers have moved closer to clinicians in their judgments (in spite of the considerable difference in procedure from Mitchell). Ullman reports, however, that these correlations may be too high by virtue of some of his procedures.

At this point it appears that the differences in directions that were a part of the Wickman procedure quite clearly contributed to the differences demonstrated between teachers and mental hygienists. When this is controlled, however, differences still emerge and these are of the kind originally observed.

It is also apparent that there has been a change in the direction of greater congruence between the attitudes of teachers and clinicians. That that congruence is not consistent for all levels of teachers has been made explicit in recent studies. It is likely that differences between elementary and secondary teachers have always existed vis-à-vis the matters here reviewed, but, where teacher and clinician attitudes appear to be the same, differences in meaning may still attach to the behaviors.

### Teacher "Expertness"

The specific criteria employed by teachers and clinicians in assessing maladjustment in children have been mentioned. As already indicated, most investigators have shown teachers to be most concerned with children's behaviors that are *aggressive, disruptive of school routines,* or generally reflecting *lack of interest in school activities.* In addition, teachers are, or have been, less concerned with *withdrawing* and other nonsocial behaviors. Some investigators have characterized this as indicative of a middle class value pattern; e.g., *stealing* is the teacher's consistent concern in MacClenathan's (24) study. The emphasis upon these school disrupting traits has not been unanimous, however. Peck (29) found *undesirable personality traits* to be the greatest concern of the teacher, *regressive* traits somewhat less so, and *aggressive* behaviors least. Of only moderate import were *violations of school work demands.* Clark (10) differs from the usual view, too, in concluding that teachers are actually more annoyed by children's behaviors which annoy other children than by behaviors which affect teachers themselves.

In most of the cited studies it is implied or explicitly stated that the teacher is "wrong" in reacting as she does to the problems of children. Teachers have been criticized as untutored in the scientific facts concerning child development and are thus seen as generally being incapable of assessing children's adjustment.

Stewart (42) rejects this thesis. With 184 boys and 193 girls as subjects, a comparison was made between ratings of problem students and nonproblem students. Identification was also attempted of those with and without "whole life" problems. From ratings of these youngsters she concludes that teachers are capable of distinguishing between problems as school problems or "whole life" problems. She insists that teachers possess much more insight into children's behavior than they are credited with by some investigators.

In spite of the few studies that report different patterns of teachers' attitudes, the hierarchy of attitudes seems to be quite close to Wickman's formulation. The Stewart report is important not so much because it rejects this hierarchy but rather in its highlighting the difference between clinicians and teachers as not being a matter of ignorance. What the difference is attributable to remains to be discussed.

### Sex Differences

There seems to be universal agreement that boys are more likely to be identified as maladjusted or behavior problems than girls (6, 8, 15, 16, 18, 20, 25, 27, 29, 32, 37, 51, 53, 55). The proportion of boys (in contrast to girls) so identified ranges in these reports from 66 to 88 per cent. Not only is there a difference in proportion but behaviors which form the bases for these identifications are in part different for each sex (*5, 15, 51*). Ullmann's (*47*) interpretation is of some interest. It is *not,* he says, that "desirable conduct for teachers takes the distinguishing characteristics of girl behavior as suggested by Wickman but rather teachers assign girls more favorable ratings because they lack awareness of the manner in which girls are making their adjustment" (p. 39). Ullmann explains that boys' patterns of adjustment are more manifest to the observer, whereas girls deal with problems on an intrapsychic level. This interpretation is in the tradition of imputing lack of insight to teachers. Stewart's (*42*) results are again cogent. Her data do not suggest that teachers lack insight into their adjustment, but rather that they distinguish a different *kind* of adjustment for girls. Another study with young adults suggests the same (*5*). In this instance, sex differences in degree of adjustment are supplemented by differences in the types of behavior identified with the maladjustments of each sex.

Why should the nature of adjustment be different for boys and girls? Whatever the *ultimate* reasons (whether biological or social), the temptation is to say that the differences, in an *immediate* sense at least, result from different *expectations.* It is evident from the cited studies that boys and girls are expected to act in prescribed ways in our culture. The reasons girls are considered better adjusted by teachers is that teachers have certain expectations of what good adjustment *in school* should be and the prescription for girls' adjustment is more consistent with these expectations than the prescription for boys' good adjustment. As Wickman makes evident, the teacher is concerned with getting what she is teaching "across," and behaviors which facilitate this are more likely to be valued. The behaviors of girls are of this kind.

This approach is more acceptable to us, from the evidence, than the interpretation that teachers' attitudes are based on a lack of sophistication.

There is some evidence that men and women teachers evaluate the problems of children differently. Women are found by one investigator to rate problem behaviors as more serious than do men (*14*). Another study (*19*) reports, however, that specific problem behaviors are treated differently by each sex. Men teachers consider *sex* problems as less serious than do women; women consider *appearance* and *destruction of property* as less serious than do men. Others (*4, 41*) report similar findings, although in the former case it is emphasized that the similarities are greater than the differences.

## Age and Grade Influences

The sixth grade appears to be modal for the nomination of children with problems, with the first and second grades offering the least. The fifth, seventh, and eighth grades also give the teacher some difficulty (*18, 20, 25, 37*). Difference in maladjustments of elementary and secondary school youngsters were reported early in the literature (*18*) and somewhat neglected till recently. Hildreth (*18*) observed that maladjusted elementary school children are more likely to be identified as *unstable, nervous,* or *shy;* the secondary school pupil, as *aggressive* or demonstrating *poor study habits.* Peck (*29*) finds the differential effect of sex of students in these identifications, however. Grade differences in problem type are reported by others as well (*5, 40, 41*). The differences found by Stouffer (*40*) have already been described. Griffiths (*16*) states that certain behavior difficulties (as reported by teachers and parents) increase with age; others decrease.

### Socioeconomic Status

There are limited data relating socioeconomic status of the child to his identification as a problem. Levy (*23*) finds "socially high grade children have personality or emotional problems . . . children of lower classes have social problems" (p. 158). Yourman (*54*) reports a larger proportion of problem children are of lower socioeconomic status. Snyder (*37*) found that schools differentiated by their level of socioeconomic status yielded different numbers of problems with more from the lower groups. There was no statistically significant difference, however, in socioeconomic status between a problem group and a control group.

The study by Griffiths (*16*) makes the most ambitious attempt to relate socioeconomic status to the identification of behavior problems. There were few significant differences among children of different socioeconomic levels in teachers' ratings of their problems. More differences appear, however, according to the parent's ratings and the child's own ratings. Griffiths concludes that some differences exist between middle socioeconomic level children and others. In particular, they are more submissive and less aggressive.

It is apparent that few data are available as to the relationship between socioeconomic status and the behavior problems of children.

## Discussion

The studies reviewed suggest strongly that differences in teachers' and clinicians' attitudes existed in 1927. From that time to the present changes appear to have taken place among teachers' attitudes so that they approximate more closely those of clinicians. There is some possibility that clinicians have tempered their evaluations as well.

Despite the shift toward congruence, teachers' attitudes remain different, and different in ways not dissimilar from what they were in Wickman's day. Why? First, let us recall that Wickman and others made much of this difference. It was suggested, even insisted, that the teacher should change. Such an injunction could rest only on the premise that the clinician's attitudes were more legitimate or more correct. This view was accepted, though G. Watson was the first and not the last to question it. Watson's position was not that the clinician was necessarily incorrect or that he should not serve as a criterion. Rather, Wickman was chided for not even considering the alternatives to accepting the clinician as criterion. For Wickman, the virtue of choosing the clinicians' attitudes as an ideal was recommended by his expert knowledge of children's adjustment. Let us examine this claim.

For one, Wickman asked clinicians to rate behaviors in light of their possible future consequences. Would the clinician, with any validity, know the future consequences of the appearance of a behavior in childhood? The answer is doubtful. In 1927 there were few if any studies which had indicated with even low degrees of certainty the outcome in adolescence or adulthood of a child's particular behavior (e.g., *withdrawing* behavior). In fact, if anything, there was some doubt that this could be done (30, 50). Evidence since then leads to even greater uncertainty (2, 21). According to some theories a withdrawn child is more likely to become maladjusted than one who is not. Yet the proof of how true this is and in what proportion for any population is almost nonexistent. Although it has been shown that in an adult *psychotic* group (52) there was a tendency for maladjustive behaviors to be present in childhood (information was obtained from retrospective reports), this tells us little about the prevalence of withdrawing behavior in a population of children or about the likelihood of such behaviors resulting in maladjustment, neurosis, or psychosis in adulthood.

However, there is a more important issue, in light of the functions of the therapist qua therapist: withdrawing behaviors present a problem to be dealt with at the time of their appearance. Such behaviors can be a basis for a visit to a therapist—in childhood as well as adulthood. The clinician is more likely to attempt some therapy than to postpone action till adolescence or adulthood (although, in some instances, this might reasonably be done). The clinician is often forced to act by immediate criteria; for he cannot wait for ultimate validation. It is thus part of his role as a therapist to be concerned about these behaviors. In essence, the behaviors with which a clinician is concerned are related to his status and the functions that accompany that status. If these behaviors were of equal relevance to the functions of the teacher, they would be equally valued. However, they are not. This has been so even in the period of "life-adjustment" programs and through the era of the "whole child," except possibly for some special groups of teachers. In spite of much pressure, teachers on the whole continue to be concerned with behaviors that facilitate or interfere with their teaching. A number of investigators recognize the difference in function between clinician and teacher (*12, 18, 24, 35, 42, 50, 54*) even though the teacher's role is not simple to define. It is, after all, a reflection of an educational philosophy. The prevailing philosophy of education in 1927, whether explicit or implicit, was oriented to the training of intellectual skills. In the interim the function of the teacher has broadened considerably to include training in social and other skills. There has been much pressure on the teacher to be a counselor and in some ways something of a psychotherapist as well—but, at the least, to focus more on the emotional life and adjustment of the child. The question of which role is "better" is a question of values. At present, the trend is back again toward the training of intellectual skills. The trend of increasing teacher sophistication in psychology will probably continue, and will probably not revert to the level of 1927. To urge (e.g., *39, 40*) that the teacher's attitudes approximate the clinician's is unrealistic unless the teacher's role becomes one with the clinician's—and this seems unlikely.

Other considerations recommend themselves as well. The teacher has a vital role in the socialization of the child. She is, after all, a culture carrier and to some extent a parental surrogate. Her own behaviors are significant in the child's development of self-control, character traits, values, and work habits. These functions are certainly as important as any. There is no question that the teacher needs to be aware of withdrawing and other undesirable personality characteristics. What is questioned is the need for the teacher to concern herself with them to the same extent and in the same way as the clinician.

To summarize, the difference reported by Wickman in attitudes toward the behavior problems of children should be interpreted as reflecting differences in the roles of teachers and clinicians and the discharge of functions of the role incumbents (*9, 34*). The efforts of many have been directed to alter the prescription of the teacher's role and performance in this role. This effort has in part been successful, as witnessed by the greater congruence in attitudes between teacher and clinician. In spite of the partial change in prescription, the teacher's role remains principally task-oriented; the clinician's, more adjustment-oriented. It seems unrealistic and possibly even undesirable to expect the teacher's behaviors reflected in her attitudes and values to become congruent with those of clinicians. Other results reviewed here are consistent with this thesis. The reported disparity between elementary and secondary teachers results from differences in role. The high school teacher is even more subject matter-oriented than the elementary school teacher. The greatest impact of the "child-oriented" or "life-adjustment" philosophy in turn has been in the elementary school. This has resulted in a modification of role prescription for the elementary school teacher which is reflected in greater similarity between the attitudinal hierarchies of elementary teachers and clinicians. The observed differences are due not only to the teacher's role but result from the actions of the children themselves. The pupil's role-related behaviors change with progress through school as the youngster assumes new responsibilities and loses old ones. Behavior differences are not only bound to their age but also their sex. The same behavior is not expected or demanded of boys and girls. The attitudes of teachers in turn will reflect differences in age and sex role expectations.

## Summary and Conclusions

The studies concerning teachers' and clinicians' attitudes toward the behavior problems of children, which have emerged principally from the initiative of the Wickman 1928 monograph, are reviewed. The following conclusions are drawn:

1. Differences existed in 1927 between the attitudes of teachers and clinicians toward the behavior problems of children. This seems to have been true in spite of the methodological limitations of the Wickman study.
2. Since 1927 there has been a shift in the hierarchy of teachers' attitudes to approximate more closely those of clinicians. This shift is not due to an artifact of research methodology. Those studies which incorporate adequate controls and consistent instructions show even greater congruence between the attitudes of the two groups.
3. There has been some change in the attitudes of clinicians although this is based upon the conclusions of one study.
4. Criteria employed in evaluating the behavior problems of children differ for elementary and secondary school teachers.
5. More boys are identified as maladjusted than girls and the criteria of maladjustment (and adjustment) differ in part for each sex.
6. The sex of the teacher affects, in part, attitudes toward children's problems.
7. Studies of the relationship of socioeconomic factors to the evaluation of children's behavior problems are inadequately dealt with in the literature.

Differences in attitudes between teachers and clinicians are interpreted in the framework of role theory. The attitudinal hierarchies of teachers and clinicians are seen as reflecting their respective roles and the ways these roles influence the organization of their respective experiences. Wickman's findings of 1927 are interpreted as indicative of the role of the teacher in that era. The role expectations of teachers have changed. Replications of the Wickman study indicate these changes have resulted in greater congruence between teachers' and clinicians' attitudes. It is suggested by virtue of the teachers' essential task-orientation and the clinicians' adjustment-orientation that complete or nearly complete congruence is not likely to be achieved.

The relationship of sex and age to attitudes toward behavior problems is also explained in the light of role theory.

## REFERENCES

1. Bain, W. E. A study of the attitudes of teachers toward behavior problems. *Child Develpm.*, 1934, **5**, 19–35.
2. Beilin, H. The prediction of adjustment over a four year interval. *J. clin. Psychol.*, 1957, **13**, 270–274.
3. Beilin, H. Effects of social (occupational) role and age upon the criteria of mental health. *J. soc. Psychol.*, 1958, **48**, 247–256.
4. Beilin, H., & Werner, E. Sex differences among teachers in the use of criteria of adjustment. *J. educ. Psychol.*, 1957, **48**, 426–436.
5. Beilin, H., & Werner, E. Sex role expectations and the criteria of social adjustment for young adults. *J. clin. Psychol.*, 1957, **13**, 341–343.
6. Blanchard, P., & Paynter, R. H. The problem child. *Ment. Hyg.*, 1924, **8**, 26–54.
7. Bowles, H. H. A study of nurses' attitudes toward behavior problems in children under hospital care. *Child Develpm.*, 1937, **8**, 282–288.
8. Boynton, P. L., & McGaw, B. H. The characteristics of problem children. *J. juv. Res.*, 1934, **18**, 215–222.
9. Brim, O. G. The parent-child relation as a social system: 1. Parent and child roles. *Child Develpm.*, 1957, **28**, 343–364.
10. Clark, E. J. Teacher reactions toward objectionable pupil behavior. *Elem. Sch. J.*, 1951, **51**, 446–449.
11. Del Solar, C. *Parents and teachers view the child; a comparative study of parents' and teachers' appraisals of children.* New York: Bureau of Publications, Teachers College, Columbia Univer., 1949.
12. Davis, E. A., & McGinnis, E. *Parent education; a survey of the Minnesota program.* Minneapolis: Univer. of Minnesota Press, 1939.
13. Dickson, V. E. Behavior difficulties that baffle teachers. *J. juv. Res.*, 1932, **16**, 93–101.
14. Ellis, D. B., & Miller, L. W. Teachers' attitudes and child behavior problems. *J. educ. Psychol.*, 1936, **27**, 501–511.
15. Epstein, L. J. An analysis of teachers' judgments of problem children. *J. genet. Psychol.*, 1941, **59**, 101–107.
16. Griffiths, W. *Behavior difficulties of children as perceived and judged by parents, teachers and children themselves.* Minneapolis: Univer. of Minnesota Press, 1952.
17. Hertzman, J. High school mental hygiene survey. *Amer. J. Orthopsychiat.*, 1948, **18**, 238–256.
18. Hildreth, G. A survey of problem pupils. *J. educ. Res.*, 1928, **18**, 1–14.
19. Hunter, E. C. Changes in teachers' attitudes toward children's behavior over the last thirty years. *Ment. Hyg.*, 1957, **41**, 3–11.

20. Hurlock, E. B., & McDonald, L. C. Undesirable behavior traits in junior high school students. *Child Develpm.*, 1934, **5**, 278–290.
21. Ives, O. L. *A critique of teachers' ratings of high school boys as an indication of later neuropsychiatric rejection for the armed services.* New York: Bureau of Publications, Teachers College, Columbia Univer., 1949.
22. Laycock, S. R. Teachers' reactions to maladjustments of school children. *Brit. J. educ. Psychol.*, 1934, **4**, 11–29.
23. Levy, J. A quantitative study of the relationship between intelligence and economic status as factors in the etiology of children's behavior problems. *Amer. J. Orthopsychiat.*, 1931, **1**, 152–162.
24. MacClenathan, R. H. Teachers and parents study children's behaviors. *J. educ. Sociol.*, 1934, **7**, 325–333.
25. McClure, W. E. Characteristics of problem children based on judgments of teachers. *J. juv. Res.*, 1929, **13**, 124–140.
26. Mitchell, J. C. A study of teachers' and mental hygienists' ratings of certain behavior problems of children. *J. educ. Res.*, 1942, **36**, 292–307.
27. Neumeyer, M. H. *Juvenile delinquency in modern society.* New York: D. Van Nostrand, 1949.
28. O'Malley, K. E. *A psychological study of the annoyances or irritations of teachers. Unpublished doctoral dissertation,* New York Univer., 1936.
29. Peck, L. Teachers' reports of the problems of unadjusted school children. *J. educ. Psychol.*, 1935, **26**, 123–138.
30. Preston, G. H., & Shepler, W. McL. A study of the problems of "normal" children. *Amer. J. Orthopsychiat.*, 1931, **1**, 245–256.
31. Pullias, E. V. How do you behave when the children misbehave? *Childh. Educ.*, 1934, **10**, 230–237.
32. Rogers, C. R. The criteria used in a study of mental health problems. *Educ. res. Bull.*, 1942, **21**, 29–40.
33. Rogers, C. R. Mental health findings in three elementary schools. *Educ. res. Bull.*, 1942, **21**, 69–79.
34. Sarbin, T. R. *Role theory.* In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Mass.: Addison-Wesley, 1954. Pp. 223–258.
35. Schrupp, M. H., & Gjerde, C. M. Teacher growth in attitudes toward behavior problems of children. *J. educ. Psychol.*, 1953, **44**, 203–214.
36. Seidman, J. M., & Knapp, L. B. Teacher likes and dislikes of student behavior and student perceptions of these attitudes. *J. educ. Res.*, 1953, **47**, 143–149.
37. Snyder, L. M. The problem child in the Jersey City elementary schools. *J. educ. Sociol.*, 1934, **7**, 343–352.
38. Sparks, J. N. Teachers' attitudes toward the behavior problems of children. *J. educ. Psychol.*, 1952, **43**, 284–291.
39. Stouffer, G. A. W., Jr. Behavior problems of children as viewed by teachers and mental hygienists. *Ment. Hyg.*, 1952, **36**, 271–285.
40. Stouffer, G. A. W., Jr. The attitudes of secondary school teachers toward certain behavior problems of children. *Sch. Rev.*, 1956, **64**, 358–362.
41. Stouffer, G. A. W., Jr., & Owens, J. Behavior problems of children as identified

by today's teachers and compared with those reported by E. K. Wickman. *J. educ. Res.*, 1955, 48, 321-331.

42. Stewart, N. Teacher's concepts of "behavior problems." In *Growing points in educational research*. Washington: American Educ. Res. Ass. Rep., 1949.
43. Stogdill, R. N. Parental attitudes and mental hygiene standards. *Ment. Hyg.*, 1931, 15, 813–827.
44. Stogdill, R. N. Experiments in the measurement of attitudes toward children, 1899–1935. *Child Develpm.*, 1936, 7, 31–36.
45. Thompson, C. E. The attitudes of various groups toward behavior problems of children. *J. abnorm. soc. Psychol.*, 1940, 35, 120–125.
46. Uger, C. The relationship of teachers' attitudes to children's problem behavior. *Sch. & Soc.*, 1938, 47, 246–248.
47. Ullmann, C. A. *Identification of maladjusted school children*. Public Hlth Monogr. No. 7. Washington: U.S. Govt. Printing Office, 1952. (Rev. 1957)
48. Wandt, E. Measurement and analysis of teachers' attitudes. *Calif. J. educ. Res.*, 1952, 3, 10–13.
49. Wandt, E. A comparison of the attitudes of contrasting groups of teachers. *Educ. psychol. Measmt.*, 1954, 14, 418–422.
50. Watson, G. A critical note on two attitude scales. *Ment. Hyg.*, 1933, 17, 59–64.
51. Wickman, E. K. Children's behavior and teachers' attitudes. New York: Commonwealth Fund, 1928.
52. Witmer, H. L. Childhood personality and parent-child relationships of dementia praecox and manic depressive patients. *Smith Coll. Stud. Soc. Wk.*, 1934, 4, 287–378.
53. Young-Masten, I. Behavior problems of elementary school children: a descriptive and comparative study. *Genet. Psychol. Monogr.*, 1938, 20, 123–181.
54. Yourman, J. Children identified by their teachers as problems. *J. educ. Sociol.*, 1932, 5, 334–343.
55. U.S. Children's Bureau. *Juvenile court statistics 1946–49*. Statistical Series No. 8. Washington: U.S. Govt. Printing Office, 1949.

# 15

# The Comparative Influence of Punitive and Nonpunitive Teachers upon Children's Concepts of School Misconduct — *Jacob S. Kounin and Paul V. Gump*

*The investigators of the research reported here were able to identify a punitive and a nonpunitive teacher at the first-grade level in each of three different schools. Individual interviews were administered pupils in these classes to determine their concepts of serious misbehavior both in school and at home. From projective interpretations (assuming that the children's responses in the interview actually did reflect their preoccupations with school conduct) the authors infer the comparative impact of punitive and nonpunitive teachers. In order to explain some of the findings, the theory is presented that children with punitive teachers develop less trust in school than do children with nonpunitive teachers.*

THIS PAPER reports a portion of a research project pertaining to the management of children's behavior in classroom settings. Because so many teachers, especially beginners, verbalize considerable concern about discipline and control, we are focusing our current research in this area.

Reprinted from *Journal of Educational Psychology,* 1961, **52,** 44–49, by permission of the senior author and the American Psychological Association.

While there is some relevant literature, such as that of Sheviakov and Redl (1944), based upon experience and insightfulness, we have been unable to locate any generalizations based upon data from research.

In a previous study by Kounin and Gump (1958) specimen-record types of observations were gathered of discipline incidents during the first week of kindergarten, focusing upon the triad of: a misbehaving child (target), a teacher doing something to stop the misbehavior, and a watching audience-child. Limiting our dependent variables to overt behavior we found that teachers' techniques of handling a misbehaving kindergarten child (target) did have different degrees of socializing success upon audience-children. A socializing success was defined as an observable reduction of overt misbehavior or an increase in conforming behavior (standing up "even straighter" in line). Control techniques high in clarity (defining the deviancy, specifying how to stop) were most successful. Control techniques high in firmness (standing closer to the misbehaving child, continuing to look at him until he stopped misbehaving) were successful only for audience-children who were themselves deviancy-oriented at the time. Control techniques high in roughness (anger, physical handling) were least successful and tended to be followed by behavior disruption (less involvement in work, overt signs of anxiety) rather than conformity on the part of audience-children. In terms of their effects, it is evident that roughness is a different dimension than firmness.

Since attitudes toward misconduct may also be affected by differences in control techniques we decided to study these as well. In an unpublished study of children at camp, P. Gump, B. Biddle, and J. Kounin found significant differences in attitudes toward camp misconduct held by campers who had effective counsellors as compared to campers who had ineffective counsellors. The counsellors, however, varied along many dimensions including punitiveness, goal-directedness, physical and psychological absenteeism, and others. The campers' attitudes toward misconduct also varied according to whether they were talking about camp, home, or school milieus. We decided, therefore, to limit the leadership dimension to punitiveness and the milieu to school.

It is postulated that aggression leads to counteraggression; it is further postulated that a punitive teacher has more power over her pupils than they have over her and that she blocks overt manifestation of pupils' aggression (observations in the classrooms of the punitive teachers selected for this study indicate that this second assumption is tenable). From these two postulations, we derive the following hypotheses:

1. That the school misconduct preoccupations of children with punitive teachers will contain more aggression than those of children with nonpunitive teachers.

2. That children with punitive teachers will be more conflicted about school misconduct than will children with nonpunitive teachers.
3. That the aggression needs and the conflict relating to misconduct hypothesized to exist among children with punitive teachers will detract from their concern with school-unique values that are not directly related to misconduct.
4. The question may also be raised as to whether or not the amount of tension generated in the children with these particular punitive teachers is sufficiently great to reduce the rational qualities of their attitudes toward misconduct.

## Method

*Subjects.* The subjects were 74 boys and 100 girls attending their first semester of the first grade in the public schools of a large city. They represented all the children from six home rooms of three schools, in from upper-lower to middle-middle socioeconomic neighborhoods.

*Procedure.* Overall school climate was controlled by selecting pairs of punitive vs. nonpunitive teachers from the same school. Three such pairs were obtained from three elementary schools.

The initial selection of punitive and nonpunitive teachers was obtained from the principal and assistant principal. Following this the classes were observed by both principal investigators. At approximately a week later the teachers were further rated by a supervisor of student teachers who visited each class twice.

The raters checked along a continuum from Extremely Punitive (threatens children with consequences that really hurt; makes threats that imply sharp dislike, real willingness to harm child; ever-readiness to punish) to Not Punitive (does not punish and does not threaten). A punitive vs. nonpunitive pair of teachers was used for the study only when all five persons agreed on their dichotomizations. All the teachers were rated as having good organization, well-behaved classes, and as achieving the learning objectives for their grade. Eighty-four of the children were in classes with punitive teachers and 90 children were in classes with nonpunitive teachers.

The children were interviewed individually during the third month of attendance at school. The interview consisted of the questions: "What is the worst thing a child can do at school?" and, following the reply, "Why is that so bad?" Identical questions were asked regarding home as the milieu for misbehavior.

*Coding the Replies.* The misconducts mentioned by the children were coded for content and for certain qualities or dimensions.

The content code (obtained from the question of "What is the worst thing to do?") contained two parts: the misconducts and the explanations given for why these were bad. The misconduct included: the act type (physical or psychological assaults, noncompliance, etc.) and the object of the misconduct (parents or teachers, other children, institutional laws or custom, etc.).

The code for the explanation of misconduct was designed to answer three questions: Who is involved in the consequence (the child himself, parent or teacher, a peer, etc.)? What kind of sufferings result to others from the misconduct (physical pain, achievement loss, property loss, etc.)? What kinds of retributions occur to the misbehaver (work imposal, character loss, physical punishment, etc.)?

## Results and Conclusions

Children probably answer the question of "What's the worst thing a child can do in school?" with a report of acts that reflect their preoccupations. It is not likely that our subjects' answers would have been the same if they were presented with a forced choice of alternative acts. Given a choice, most children would probably rate "stabbing someone" as more serious than "talking in class." If the misconducts the children talked about are taken to represent tension systems and preoccupations, we may infer from these the comparative impact of punitive and nonpunitive teachers.

In a concurrent study of children's attitudes toward misconduct (201 boys and 214 girls in the first grade of six public schools representing a range of socioeconomic backgrounds). Gump and Kounin (1959) found both sex differences and differences between home and school milieus. For example, home misconducts included more breaking of objects while school misconducts included more rule violations; parents suffered more than teachers in consequences but teachers retributed more frequently. However, parents were reported as retributing with more corporal punishment and with more severe punishment than teachers. There were also differences in the responses of boys and girls, especially in school. For example, girls reported "talking" as a school misconduct eight times more frequently than boys, whereas boys reported physical assaults on peers in school more frequently than did girls.

Consequently, the comparison of the responses of children with punitive and nonpunitive teachers was made separately for sexes and also for home and school milieus. However, on *all* comparisons of school responses the direction of differences between children with punitive and nonpunitive teachers was the same for both boys and girls. The report of results, therefore, combines both boys and girls. Insofar as the differences between children with punitive and nonpunitive teachers are concerned, only 2 of the 48 comparisons of home responses were statistically significant: home misconducts of the children with punitive teachers were rated as more serious

($p < .05$, for girls only) and retributions to the subject were more serious ($p < .02$, for boys only). It is uncertain whether these represent some spillover of the influence of punitive teachers onto attitudes toward home misconducts, or whether they are chance differences for the number of comparisons made.

The results to be reported here, then, refer to boys and girls combined and to school misconducts only. Intercoder reliabilities ranged from 73–95% agreement, with a median of 90. The $p$ levels of differences are based on the $\chi^2$ test. In the case of dimensions, such as "seriousness," the results were dichotomized into a High and a Low based upon as equal a break as was possible and resulting $\chi^2$'s were based upon $2 \times 2$ tables. In the case of categories such as act-types falling into the categories of: rule violations, physical assaults on children, property damages, or nonconformance with adults, the $\chi^2$ was computed for as many cells as there were categories. At times, when one particular category was of interest, a $2 \times 2$ table was constructed with that category v. "all the rest," providing that the overall table showed statistical significance.

Following are definitions of the codes used which are not self-explanatory both for the misconducts and for the subjects' explanations for the wrongness of the act. These codes appear in Table 1.

TABLE 1. A Comparison of Attitudes toward School Misconducts Held by Children with Punitive and Nonpunitive First Grade Teachers ($N = 176$)

| Misconducts and Explanations | % of Pu[a] | % of NPu[a] |
|---|---|---|
| I. Content and quality of the misconducts: | | |
| A. Physical assaults on others | 38 | 17 |
| B. Milieu-serious misconducts | 89 | 63 |
| C. Coder-serious misconducts | 48 | 27 |
| D. Abstract misconducts | 27 | 52 |
| II. Content and quality of the explanations: | | |
| A. Peers as objects of consequences | 94 | 61 |
| B. Physical damage to objects of consequences | 60 | 23 |
| C. Serious harm to others | 45 | 18 |
| D. Reality-centered retributions | 21 | 48 |
| E. "Reflexive justifications" as explanations | 11 | 26 |
| III. Role of self in misconducts: | | |
| A. Ego-alien misconducts | 26 | 11 |
| B. Premeditated misconducts | 29 | 15 |
| IV. Aggression: | | |
| A. Overall aggression ("blood and guts") | 49 | 24 |
| V. Concern with school-unique objectives: | | |
| A. Learning and achievement losses | 20 | 43 |
| B. Institutional law violations | 49 | 62 |

Note: All differences in percentages are significant at the .05 level or beyond.

[a] Pu stands for those children who have punitive teachers; NPu refers to those children who have nonpunitive teachers.

### I.

A. Physical assaults include all physical attacks on other persons (pushing, hitting).

B. Milieu-seriousness refers to the length to which the milieu would go to prevent such an act. The school would practically ignore "scratching head," would mildly frown at "whispering," and would go to any length to stop burning down buildings.

C. Coder seriousness refers to the general immorality or danger in the misconduct considered from the point of view of the overall culture. (The only frame of reference which produced high intercoder agreement was when the coders took the position of an understanding Reformed rabbi or a Unitarian minister. Taking an unspecified role, or that of either a parent or a teacher produced low intercoder agreement.) "Studying spelling lessons at the wrong time" is morally trivial while "maiming someone" is morally very serious.

D. Abstractness of misconduct refers to the size of coverage. It may range from a unique, "one time" misbehavior, such as, "cut a climbing rope in gym," to an abstract one, such as, "be mean to other people."

### II.

A. A central adult is the responsible leader: teacher at school.

B. A psychological loss to another is exemplified by "It would make her worry."

C. Seriousness of consequences to others range from trivial harm, such as, "She'd be annoyed," to serious ones, such as, "He'd die."

D. A reality-centered retribution (this is scored only when the perpetrator himself suffers in the consequence) is coded when the consequence of a misconduct follows naturally from the act-type, such as: "not study because you'll get behind in your work." This contrasts with the response in which the connection between act and consequence is dependent upon a personal intervention of another, such as: "not study because teacher will make you stand in the corner."

E. "Reflexive justification" was coded when the child gave no consequence for either himself or others in his explanation of why the act was bad. When he said the act is bad because "It's not nice" or "It's bad" it was called a reflexive justification.

### III.

A. On ego-acceptability, we sought to determine the degree to which the respondent could see himself as the perpetrator of the misconduct. In an ego-alien act, the respondent expresses abhorrence, such as: "It's dirty to hit little kids who didn't do nothing to you." An ego-attractive act is one in which the child indicates its seductive quality for him, such as: "Tell off a teacher—boy, I'd like to do that."

B. On the premeditation category, we sought to learn the extent to which the child sought to do wrong. If premeditated, the child plans the act and intends the consequences ahead of time, such as: "Put thumb tacks on teacher's chair when she is out." If intentional, the child accepts his part in the wrongdoing but does not plan it, such as, "talk during a lesson."

IV.

A. Aggression ("blood and guts") refers to the amount of aggression the respondent expresses in his misconducts and consequences. "Play in the storage bin because somebody might get hurt" expresses less aggression than "Play in the storage bin because you might push a kid off and there could be a sharp rock down there and he could hit his head against it and crash open his skull and he would bleed and his brains would fall out and he'd die."

V.

A. A learning or achievement loss is coded when interference with learning is the misconduct or the explanation, such as: "It's bad to make noise because somebody could make a mistake in his work," or ". . . because then he couldn't read good."

B. An institutional law violation is a violation of the rules of the school such as: "talk when you're supposed to study," "not take your seat when the bell rings."

The results presented in Table 1 may be summarized around the three hypotheses and the one question raised in the introduction:

*Punitive teachers will create or activate more aggression-tension than will nonpunitive teachers.* This is strongly supported by the data. The children who have punitive teachers have more sheer aggression in their sins and consequences, they give both more milieu-serious and more coder-serious misconducts, their targets suffer more harm, they give more physical assaults as act-types, and their targets suffer more physical harm. The targets of children with nonpunitive teachers are more inclined to suffer psychological losses as consequences. As an example of the results: of 84 respondents with punitive teachers, 31 give physical assaults on other children and 40 mention school rule violations; while of 90 children with nonpunitive teachers, 15 talk about physical assaults and 56 about rule violations. (The remainder of the act-types are nonconformances and "miscellaneous.")

*Children with punitive teachers will be more unsettled and conflicted about misbehavior in school.* This hypothesis is supported by the findings related to the role of self in misbehavior. The children from nonpunitive teachers give misconducts in which their own role is intentional whereas children from punitive teachers give both premeditated and ego-alien misconducts. We may say that children with punitive teachers express more abhorrence for the misdeeds which they have selected and yet select misdeeds which require "malice and forethought."

*Punitiveness of teachers will detract from children's concern with school-unique values.* This hypothesis is supported. Children from punitive teachers talk more about physical attacks on peers—misbehavior by no means unique to the classroom setting. Children with nonpunitive teachers talk more about learning, achievement losses, and violations of school-unique values and rules.

*Do children from nonpunitive teachers show more rational qualities in their responses?* The answer to this question is not clear. Fairly direct attempts to measure this—codes for milieu likelihood of misconducts, for likelihood of consequences, and for appropriateness of consequences to the misconduct—did *not* show significant differences between the two groups. On the other hand, children with punitive teachers gave fewer abstract misconducts which result, in our camp study, was negatively correlated with age. But these same children also gave fewer reflexive justifications which result was positively correlated with chronological age. One interpretation of the findings that children of punitive teachers gave fewer abstract misconducts and fewer reflexive justifications is to regard these as indications of the unsettled and conflicted state of the attitudes regarding misconduct held by children with punitive teachers. When a child is inclined to misbehave but fears to, then a concrete act occurs to him—"hit George in the mouth"; when he is not pressed by his needs to misbehave, then an abstraction occurs to him—"be mean to people." Similarly, when he expresses this verbal act, a real consequence occurs—the target gets hurt or the perpetrator suffers a consequence; when he is not preoccupied with wrongdoing then a reflexive justification occurs to him—"it's not nice." A reflexive justification at this age may not be a primitive reply but a reflection of a settled issue: "You just don't do this because it's not nice."

Another interpretation is to regard the greater use of reflexive justification by the children with nonpunitive teachers as evidence of their greater trust and faith in school, i.e., of their internalization of school values more than children with punitive teachers. Inspection of the data showed the reflexive justification was used predominantly in connection with rule violations (talking, running in halls, not taking seat, and the like). These misconducts are milieu-inconvenient which are disturbing to the milieu but which are without direct harm to either the actor or to others and do not violate an important moral code. Such misconducts to the first grade child have no real explanation except that "they're bad because they say so." As such, they express a sort of naive faith and trust in the rightness of what the teacher says.

## Summary

Three pairs of punitive vs. nonpunitive first grade teachers were selected from three elementary schools. The 174 children in these teachers' classrooms were individually interviewed about what they thought was "the worst thing to do in school" together with their explanations of why these misconducts were bad. Regarding their responses as expressions of their preoccupations it was concluded that, as compared with children who have nonpunitive teachers, children who have punitive teachers: manifest more aggression in their misconducts, are more unsettled and conflicted about misconduct in school, are less concerned with learning and school-unique values, show some, but not consistent, indication of a reduction in rationality pertaining to school misconduct. A theory that children with punitive teachers develop less trust of school than do children with nonpunitive teachers was also presented to explain some of the findings.

## REFERENCES

Gump, P. V., & Kounin, J. S. Sex and milieu influences in children's concepts of misconduct. *Amer. Psychologist*, 1959, 14, 346. (*Abstract*)

Kounin, J. S., & Gump, P. V. The ripple effect in discipline. *Elem. sch. J.*, 1958, **59**, 158–162.

Sheviakov, G., & Redl, F. Discipline for today's children and youth. Washington, D.C.: National Education Association, 1944.

# 16

# Relationships Between Pupil Achievement, Pupil Affect-Need, Teacher Warmth, and Teacher Permissiveness — *C. M. Christensen*

*Many workers in the field of education look upon pupil performance as being influenced at least in part by motivational factors related to teacher behavior. The present author reports an exploratory study of the relationship between permissiveness and warmth of teachers, and affect-need and achievement of pupils. Teacher warmth and permissiveness are assessed by obtaining pupil ratings. (Because pupils may ignore behaviors that cutside observers assume are influential, in this type of study ratings based on pupil perception may be more valid than ratings based on outside observation.) A cognitive-affective scale was used to assess pupil affect-need. Warmth of teachers was significantly related to students' vocabulary and arithmetic achievements—a finding interpreted as indicating that teacher affect may be an important motivating factor in student achievement. Although none of the other relationships examined were found to be significant, the exploratory nature of the present study precludes drawing any definitive conclusions about those relationships.*

GUIDANCE OR direction of pupil learning and general arousal or motivation of pupils are, in the present study, considered two important

Reprinted from *Journal of Educational Psychology*, 1960, **51**, 169–174, by permission of the author and the American Psychological Association.

factors influencing the achievement of pupils. The directiveness-permissiveness (permissiveness) of the teacher in teaching subject matter is considered an aspect of the guidance factor. The affective response (warmth) of the teacher to the pupil and affect-need of the pupil are considered significant aspects of motivation.

The study of teacher permissiveness and warmth as separate dimensions is likely to help clarify teaching process problems. Cronbach (1954, Ch. 15) makes a similar distinction and points out that there has been a tendency to equate warmth with permissiveness and harshness with directiveness. Casual observation suggests that it is possible for a warm teacher to be either permissive or directive. Similarly, a harsh teacher might be either permissive or directive in her teaching of subject matter. When permissiveness is considered separately from warmth, it is plausible to hypothesize that permissiveness will be negatively related to achievement of pupils.

Regarding warmth of teacher as a motivational factor, a positive response by the teacher is expected to result in optimum arousal or motivation, whereas an indifferent or negative response is expected to result in either less than or greater than optimum motivation. Literature in the area of psychotherapy could be cited to point up the general reactions of individuals to the affective response of others.

The effect of the teacher's affective response on the pupil would be modified to some extent by the needs of the pupil. For example, pupils with high affect-need would probably react differently to teacher responses than pupils with high cognitive-need (Della Piana & Gage, 1955). This, of course, would have a bearing on the effectiveness of teacher response as a motivational or arousal factor.

The present study was concerned with pupil achievement as the dependent variable and pupil affect-need, teacher warmth, and teacher permissiveness as independent variables. The following hypotheses were tested:

1. Positive affective response (warmth) of teacher is positively related to achievement gains.
2. Permissiveness of teacher is negatively related to achievement gains.
3. Teacher warmth and permissiveness interact significantly such that warm, directive teachers will produce the greatest achievement gains.
4. Affective needs of pupils interact significantly with teacher warmth and permissiveness.

## Method

### Subjects

Subjects included 10 fifth-grade classes of pupils, 10 fourth-grade classes of pupils, and 10 fourth-grade teachers. This represented all the fourth- and fifth-grade pupils, and all the fourth-grade teachers in one New York state suburban school system.

### Instruments

Brief descriptions of the instruments used in this study are given below. Some of the items included in the Warmth and Permissiveness scales were devised by the writer, others were selected from existing scales (Della Piana, 1953; Leeds, 1950; Leeds, 1954; Medley and Klein, 1956).

*Permissiveness.* The Permissiveness scale attempts to measure the permissive-directive behavior of the teacher in her teaching of subject matter. Forty of the following kinds of items were included:

Do the pupils usually help plan what the class is going to do?
When you work arithmetic problems do you have to show all your work?
Does your teacher make assignments every day?
Does your teacher assign the pages to be read in your science book?
Does your teacher push some pupils to try a little harder?

Pupils responded yes or no to all items. The teacher score was obtained by computing a mean of the pupil scores. High teacher scores represent greater permissiveness.

*Warmth.* The Warmth scale attempts to measure the affective response of the teacher to pupils. Forty of the following kinds of items were included:

Does your teacher ever laugh and joke with the pupils?
Is it easy to talk to your teacher when you feel bad about something?
Is your teacher easily annoyed or bothered?
If you made a mistake would you be afraid to tell your teacher about it?
Does your teacher ever say mean things to the pupils?

Like the Permissiveness scale, pupils responded yes or no to all items and the teacher score was obtained by computing a mean of the pupil scores. Higher teacher scores represent more positive affective response.

*Affect-Need.* A cognitive-affective scale[1] devised by Della Piana and Gage (1955) was used to measure affect-need of pupils. The scale consists of 36 paired-comparison items of the following kind:

Which do you want more in a teacher?
  1A. Explains so we can understand.
  1B. Is nice to us even if we do something wrong.

In the present study a high score reflects high affect-need.

*Achievement.* The Iowa Tests of Basic Skills were used to measure pupil growth in achievement. This is a standardized achievement battery including the following five subtests: (*a*) Vocabulary, (*b*) Reading Comprehension,

---

[1] Thanks is due N. L. Gage for providing a copy of the scale.

(*c*) Language Skills, (*d*) Work-Study Skills, and (*e*) Arithmetic Skills. The battery also provides one overall score called a composite score.

### Collection of Data

The Warmth, Permissiveness, and Affect-Need scales were administered to the fourth-grade pupils. All items were read to the pupils by the writer. The teacher was not present during the administration of the scales.

The Affect-Need scale was administered, in a similar fashion, to the ten classes of fifth-grade pupils. In addition the Iowa Tests of Basic Skills had been administered to these pupils, during the first month of the fourth grade and the first month of the fifth grade. This permitted the measurement of growth in achievement for the present fifth-grade pupils during their stay in the fourth grade.

It should be noted that the teacher Warmth and Permissiveness scores were derived from the fourth-grade pupil responses and growth in achievement was derived from the previous year's fourth-grade pupils who were now in the fifth grade. For the statistical analyses, the fifth-grade pupil achievement scores were grouped according to their fourth-grade teachers.

## Results

### Characteristics of Scales

Table 1 gives evidence that the teachers differ significantly on both the Warmth and Permissiveness scales. Analyses of variance yielded significantly greater among-teacher variance than within-class variances. Variances for pupils within classes is small compared to the variances computed for teacher scores based on mean pupil ratings.

TABLE 1. Analysis of Variance of Permissiveness and Warmth Scores Obtained from Ten Classes of Pupils

| | | Permissiveness | | Warmth | |
|---|---|---|---|---|---|
| Source | *df* | *MS* | *F* | *MS* | *F* |
| Among teachers | 9 | 96.00 | 15.74* | 262.55 | 11.30* |
| Within classes | 217 | 6.10 | | 23.23 | |
| Total | 226 | | | | |

* Significant at .01 level.

Following a procedure used by Mitzel and Medley (1957), reliability coefficients for teacher scores on the Warmth and Permissiveness scales were estimated from the data given in Table 1. With this procedure, among-teacher variance is used as an estimate of obtained score variance, and among-teacher variance minus within-classes variance as an estimate of true score variance. Reliability is defined as the ratio of true score variance to obtained score variance. The reliability coefficient for the Warmth scale was .91 and for the Permissiveness scale it was .94. This indicates that other classes of similar pupils would rank the teachers in almost the same order as the classes who rated the teachers in the present study.

Independence of the Warmth and Permissiveness scales was also considered. An adequate test of independence should be based on the mean pupil ratings of their teacher. The rank correlation between Warmth and Permissiveness for ten teachers was .39. Although a rank correlation of this magnitude is not significant, the number of cases is too small for a rigorous test.

Correlations, computed separately for boys and girls, between Warmth, Permissiveness, and Affect-Need are given in Table 2. These correlations were based on individual pupil responses. Low positive but significant correlations were obtained between Warmth and Permissiveness. Affect-Need did not correlate significantly with either Warmth or Permissiveness. There is no evidence of sex differences.

TABLE 2. Product Moment Correlations Between Affect-Need, Permissiveness, and Warmth Scores of 103 Boys and 121 Girls

| | Warmth | | Permissiveness | |
|---|---|---|---|---|
| | Boys | Girls | Boys | Girls |
| Affect-Need | −.17 | −.06 | .09 | .02 |
| Warmth | | | .32° | .25° |

° Significant at .01 level.

Della Piana and Gage (1955) reported a Horst reliability coefficient of .74 for the Affect-Need scale.

**Tests of Hypotheses**

The mean Warmth score and mean Permissiveness score were obtained for each of the ten teachers. Eight of the teachers with the most extreme scores were classified in four groups, two teachers to a group. The four groups resulting were as follows: High Permissiveness and High Warmth, High Warmth and Low Permissiveness, Low Warmth and High Permissiveness, and Low Warmth and Low Permissiveness. The pupils of the teachers in each of the groups were subdivided in terms of Affect-Need scores. For each group of teachers, ten high Affect-Need and ten low Affect-Need pupils were selected. This yielded a 2 × 2 × 2 factorial design with 10 individuals per cell.

Covariance analysis was used to test the hypotheses. Beginning fourth-grade achievement scores were used to adjust the beginning fifth-grade achievement scores. This adjustment subtracts the predicted beginning fifth-grade scores from the actual beginning fifth-grade scores. The scores used in the analyses can be thought of as growth in achievement.

Table 3 gives a summary of the analyses for each of the subtests and composite scores for the Iowa Tests of Basic Skills. Only two significant results were obtained. Vocabulary and arithmetic achievement growth were significantly greater for teachers scoring high on the Warmth scale. No significant relationships were obtained for Affect-Need or Permissiveness and none of the interaction terms were significant. Only the first hypothesis was partially substantiated.

TABLE 3. Analysis of Variance of Adjusted Achievement Test Scores

| | | Vocabulary | | Reading Comprehension | | Language Skills | | Work-Study Skills | | Arithmetic Skills | | Composite | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Source | *df* | *MS* | *F* | *MS* | *F* | *MS* | *F* | *MS* | *F* | *MS* | *F* | *MS* | *F* |
| Permissiveness | 1 | 9.59 | | 71.52 | | 17.19 | | 33.65 | 1.26 | 2.02 | | 5.40 | |
| Warmth | 1 | 259.44 | 4.65° | 17.78 | | 156.97 | 2.22 | 30.04 | 1.12 | 113.36 | 6.38° | 49.39 | 3.16 |
| Affect-Need | 1 | 19.65 | | 301.55 | 3.55 | 24.22 | | 5.26 | | 24.83 | 1.40 | 21.52 | 1.38 |
| P × W | 1 | .42 | | 130.36 | 1.53 | 44.54 | | .12 | | 1.42 | | .51 | |
| P × A | 1 | 100.58 | 1.80 | 13.88 | | 1.11 | | 4.87 | | .31 | | 2.38 | |
| W × A | 1 | 2.71 | | 97.77 | 1.15 | 27.45 | | 52.48 | 1.97 | 25.63 | 1.44 | 25.23 | 1.61 |
| P × W × A | 1 | 32.51 | | 29.91 | | 24.46 | | 53.10 | 1.99 | 29.18 | 1.64 | 1.71 | |
| Error | 71 | 55.76 | | 85.02 | | 70.56 | | 26.70 | | 17.77 | | 15.64 | |
| Total | 78 | | | | | | | | | | | | |

° Significant at .05 level.

## Discussion

Evidence from the present study supports the notion that warmth and permissiveness can be studied as separate dimensions. The correlation between the Warmth and Permissiveness scales was low and both scales discriminated between the teachers. The high reliability coefficients obtained indicate that pupils within a class are consistent in describing their teacher. Pupils within a class can be regarded as several observers rating one individual. Assuming that errors of observations include both over- and underestimation, a score based on the mean of the ratings will provide an accurate measure.

Individual pupil ratings of teachers and Affect-Need scores of pupils were intercorrelated. Affect-Need scores did not correlate significantly with the pupil ratings of teachers. A low, positive but significant relationship between Permissiveness and Warmth was obtained. This could be interpreted either as a tendency for pupils to regard warm teachers as permissive or as an indication that warm teachers are, in fact, more permissive. The low intercorrelations may be due to low reliability of the scales. It should be noted that reliability coefficients for individual pupil ratings and Affect-Need scores were not computed.

Of the variables considered, only warmth of teacher was related to vocabulary and arithmetic achievement. Plausible explanations can be offered for this differential relationship. Arithmetic achievement is much more dependent on school factors than, for example, reading. Little direct attention is given to arithmetic in the home and few pupils make daily use of arithmetic outside of school. The relationship between vocabulary achievement and warmth is, however, not as obvious as in the case of arithmetic. A reading specialist[2] has pointed out that in attempts to improve reading, vocabulary is more resistant to improvement than reading comprehension. She suggested that vocabulary depends on interest in words which is likely to be facilitated by a responsive teacher. Additional research is needed to clarify these interesting possibilities.

Results support the contention that affective response of the teacher is more important for growth in achievement than permissiveness. This is an interesting finding, particularly if it is further substantiated by additional research. It has theoretical implications for the motivation of pupils and it also has implications for teacher training. Undoubtedly the development of affective responsiveness in teachers would call for quite different kinds of techniques than those now generally employed in teacher training programs.

Affect-Need was in no instance related to achievement. This was somewhat surprising. Della Piana and Gage (1955) reported a differential relationship between Minnesota Teacher Attitude Inventory scores and pupil-rated liking of teachers for pupils of high and low Affect-Need. Of course, liking of teacher may not be related to achievement. However, it was expected that Affect-Need and Warmth would interact and influence achievement.

A final note of caution is in order. This study should be regarded as exploratory in nature and not as definitive. It was obviously not an experiment but rather a study of relationships between selected teacher and pupil characteristics. A school system was entered intact and no manipulation of conditions took place. In contrast to an experiment, the pupils were not randomly assigned to teachers. Although some degree of control was obtained by the covariance analysis, it is likely that other pupil characteristics influencing achievement growth were operating. Random assignment of pupils to teachers is probably the only effective way of controlling unknown, relevant variables.

[2] Marion Dixon Jenkinson, University of Alberta, personal communication.

## Summary

This study explored relationships between permissiveness and warmth of teachers and affect-need and achievement of pupils. Permissiveness and Warmth scales were devised and administered to pupils in ten fourth-grade classes. An Affect-Need scale devised by Della Piana and Gage was administered to ten fifth-grade classes. Beginning fourth-grade and beginning fifth-grade achievement test scores were available for the fifth-grade pupils. The Permissiveness and Warmth scales were sufficiently reliable and independent for the purposes of this study. Affect-Need did not correlate significantly with either of the scales. A 2 × 2 × 2 factorial design with two levels of permissiveness, warmth, and affect-need was established. Covariance analysis was used to determine growth in achievement. Contrary to hypothesized outcomes, none of the interaction, permissiveness, or affect-need variances were significant. Warmth of teachers was significantly related to vocabulary and arithmetic achievement.

## REFERENCES

Cronbach, L. J. *Educational psychology.* New York: Harcourt, Brace, 1954.

Della Piana, G. M. The cognitive-affective values of pupils and teacher-pupil relationships. Unpublished master's thesis, Univer. Illinois, 1953.

Della Piana, G. M., & Gage, N. L. Pupils' values and the validity of the Minnesota Teacher Attitude Inventory. *J. educ. Psychol.*, 1955, **46**, 167–178.

Leeds, C. H. A scale for measuring teacher-pupil attitudes and teacher-pupil rapport. *Psychol. Monogr.*, 1950, **64**, No. 6 (Whole No. 312).

Leeds, C. H. Teacher behavior liked and disliked by pupils. *Educ.*, 1954, **75**, 29–37.

Medley, D. M., & Klein, A. A. *Studies of teacher behavior: Inferring classroom behavior from pupil responses.* New York: Board Higher Educ., New York City, Div. Teacher Educ., Office Res. Evaluation, 1956.

Mitzel, H. E., & Medley, D. M. Pupil growth in reading—an index of effective teaching. *J. educ. Psychol.*, 1957, **48**, 227–239.

## 17

# Group Compliance to Dominative Teacher Influence — *Ned A. Flanders and Sulo Havumaki*

*The authors of this article have done a considerable amount of research, at the University of Minnesota, on the relationship between teacher and student behavior in the classroom. In the present article they are concerned with the effects of teacher and group pressures on student opinion. They compare changes in opinion among groups of tenth-grade students under four different experimental conditions involving patterns of teacher behavior and of group responses to teacher's arguments. Under the conditions of this experiment, group conformity appeared to be a more potent force in influencing opinion than did teacher persuasion. Alongside other important findings, it was apparent that students who changed their opinions to comply with teacher and peers might have done so only on the surface; thus considerable inner resistance to teacher opinions might still exist. Several interesting implications for classroom behavior and management, and for individual student adjustment are discussed.*

### The Problem

A TEACHER sometimes faces a situation in which there is an issue that is a matter of opinion. In its simplest form, the issue may involve two alternatives, say *A* and *B*. Suppose the teacher prefers *A* and nearly all the

Reprinted from *Human Relations,* 1960, 13, 67–82, by permission of the senior author and Tavistock Publications, London.

students prefer *B*. Ideally, the ensuing social interaction tends to reduce the difference of opinion by seeking 'the best' solution. But teacher-pupil relationships are essentially superior-subordinate, and the teacher has ultimate authority. Not only does the teacher's opinion carry more weight than students' opinions, but students expect the teacher to be 'right.' In such a situation, changing opinion from *B* to *A* is an act of compliance requiring the students to give up their preferences. Holding fast to opinion *B* is to resist complying to the teacher's influence.

Given the difference of opinion, a teacher usually exerts logical persuasion, explains the problem as he sees it, outlines certain facts or reasons that support his own point of view, and frequently asks questions designed to stimulate student discussion. Modern practices notwithstanding, a teacher usually will talk much more than all the students combined and will control the communication pattern by responding to each student's statement. Soon all student comments are directed to the teacher, and even though the teacher is pleasant, respects the students, and appears to be fair, the basic superior-subordinate relationship is maintained. These are precisely the conditions that were created in our laboratory study.

Our theoretical analysis of the foregoing situation makes use of two independent variables that affect compliance. The first independent variable is the 'pattern of influence' used by the teacher. Clearly *how* the teacher (*T*)[1] uses his influence will affect student (S)[1] compliance. The second independent variable is group contagion or the tendency of a S to comply because he thinks others are complying or not to comply if he thinks others are resisting the *T*'s influence.

The pattern of teacher influence is limited, first of all, to logical, friendly argument without threat of punishment or promise of special reward other than *T* commendation. Within this limitation, two patterns are used. In one-half of the experimental groups, the *T* directed his arguments only to individuals while the rest of the group listened. In the individual treatments of our experimental design, the *T* called on only those Ss who happened to be seated in the odd-numbered seats, always referred to Ss by name, addressed his remarks to individuals, and spoke to the group as infrequently as possible. As one would expect, this behavior caused Ss to raise their hands to be called on by the *T* and in this way he was able to restrict communication with ease. In the other one-half of the experimental groups, the *T* directed his arguments to the total group. He spoke only to the group, told the group that anyone should speak right out without raising his hand or waiting to be called on, reacted to ideas expressed as if they belonged to the group, and did not single out individuals by the use of names.

---

[1] The letter *T* will refer to the teacher; the letters S or Ss will refer to the singular or plural of student; and the letter *E* will refer to the host experimenter who is in charge except during class discussions.

The variable of group contagion was manipulated[2] by the use of special equipment in the laboratory room. All experimental groups were taught to use individual levers hidden from view to indicate their preference for either alternative *A* or alternative *B*. A light panel, hidden from the *T*'s view but in full view of all Ss purported to show the total lever movements for both alternatives. Each *S* could determine the total group opinion by noting the proportion of red and green lights. In one-half of the experimental groups the light panel, which was actually controlled by the *E*, who sat in an observation room, consistently showed throughout the discussion an almost unanimous preference for alternative *B*, the original preference of the Ss. In this condition Ss perceived others as not complying to the persuasion of the *T*. In the second experimental condition the light panel gradually changed, during the discussion, from a nearly unanimous preference for *B* to a nearly unanimous preference for *A*. In this condition Ss gained the perception that the group was shifting its opinion from *B* to *A*. In this manner group perceptions essential to contagion were controlled.

## Hypotheses and Experimental Design

The first three hypotheses are concerned with differences in compliance that one might expect when the individual or the group approach is used by the *T*.

*Hypothesis 1.0.* The individual approach will result in greater compliance than the group approach.

*Sub-Hypothesis 1.1.* Group cohesiveness will be lower with the individual approach, compared with the group approach.

*Sub-Hypothesis 1.2.* Dependency on the teacher will be higher during the individual approach, compared with the group approach.

The rationale behind these hypotheses is that the direct influence attempts of the individual approach are divisive. Each contact emphasizes the *T*'s concern with the individual's opinions and disregards relationships between individuals. This laboratory procedure simulates the common classroom practice of dealing with the Ss 'one at a time'. In a situation in which the *T* is trying to persuade Ss, it is the divide-and-conquer theory. One would predict, in these circumstances, that a measure of cohesiveness, or the strength of each individual's desire to remain in the group, will be lower in the individual approach.

---

[2] Manipulation involves professional ethics. The authors feel that these procedures are justified to control experimental conditions only when (a) no individual or group suffers emotionally, and (b) no one leaves the experiment proper with any misconceptions that might adversely affect his social adjustment.

Superimposed on the divisive aspects of the individual approach is the expectation that dominative teacher influence creates dependency. Dependency differs from compliance in the following way. A direct act of persuasion, if successful, leads to compliance if the S shifts his opinion, i.e. compliance is an act that is in response to an order or to general pressure. An act of dependency is a voluntary act such that a *S* seeks approval or direction from the authority figure. In one sense, it is a voluntary seeking of additional areas in which acts of compliance can be carried out.

Lower group cohesiveness and higher dependency on the teacher will weaken the ability of the group to resist teacher influence, thus, compliance should be higher. The group approach does not restrict or limit the group resources to the same extent. Given hypotheses 1.1 and 1.2, the main hypothesis 1.0 naturally follows.

The second major hypothesis is concerned with the differences in compliance that one might expect when group contagion is manipulated.

*Hypothesis* 2.0. Compliance increases when students perceive that others are changing their opinion in response to the teacher's persuasion.

This hypothesis is based on well-established principles of conformity to group pressures. In this experiment, however, the forces of group contagion sometimes support and sometimes oppose the persuasive efforts of the *T*. Control of contagion forces by manipulating the light panel combined with the two methods of *T* influence provide four treatments in our experimental design. These four treatments are:

| *Treatment* | *Teacher Approach* | *Group's Perception* |
|---|---|---|
| G-O | Group | Opposition |
| I-O | Individual | Opposition |
| G-C | Group | Compliance |
| I-C | Individual | Compliance |

If hypotheses 1.0 and 2.0 are to be supported by the data, compliance should be greatest in the I-C treatment and smallest in the G-O treatment. The rank-order of treatment I-O and G-C, according to compliance, will depend on the relative magnitude of the force toward compliance created by the contrasting approaches of the teacher versus the relative magnitude of the contagion manipulation. If the forces associated with contagion are greater, the rank-order as listed above should occur.

### The Procedure

1. Each experimental group of exactly ten members was brought by bus from a local high school to the University laboratory rooms, where they were welcomed by the *E*. The Ss were told they had been chosen by chance to form a school team for a city-wide high-school competition that would operate like a 'Quiz Kid Show.' Five members of the team would be called 'participators' and five would be called 'helpers.' The participators would do the talking during the quiz show. The helpers would also be present in the studio, seated right behind the participators, and could coach or help the participators in case the latter did not know the answer. The Ss were told that the main reason for designating participants and helpers was that, if the twenty boys and girls from both teams were all permitted to talk, there would be too much confusion. The whole introduction, the rules for the contest, and general build-up were carefully planned to convince even the most skeptical *S* that the competition was real and serious. It was explained that each team could choose to compete on either radio or television. Several statements were made about TV shows being lots of fun and that most Ss would prefer TV. At the time, TV was quite new in Minnesota and very attractive to Ss.

2. Ballots, inconspicuously coded, were used to collect the prediscussion preferences for TV or for radio. The private ballots were collected and the Ss informed that the results would be revealed to them in a few minutes.

3. The levers and the light panel were then explained and the Ss practised using them. The demonstration was designed to convince the Ss that *their* lever movements controlled the proportion of red (TV) and green (radio) lights. Through discussion with the *E*, the Ss were persuaded that the levers could be used by them in a discussion that was about to follow to indicate their preferences continuously and automatically. The equipment would permit them to know just how the entire group felt about the two alternatives as the discussion progressed, and this would result in a more adequate group decision because everyone's opinion could be expressed at all times.

4. The Ss were then told that each team in the contest was guided by a teacher-trainer, and that the teacher-trainer (*T*) assigned to their team would now meet with them to lead their group's discussion. During the discussion they would reach a final decision about radio or TV. The Ss were again reminded of the light panel, 'You can use the light panel to see if your own opinion is the same or different than the rest of the group.' The results of the earlier secret ballot were then announced. Regardless of how the group actually voted, the results for each group were: 'According to your vote, most of the members of this group prefer TV. Therefore, when the light

panel is turned on most of the lights will be red.' At the beginning of each group discussion 12 red lights and 2 green lights appeared.

5. The *T* was then introduced to the group by the *E*. The *T* began the discussion by asking what they thought about TV versus radio as media for their group. Depending on the treatment, either an individual or a group approach was used by the *T* throughout the discussion; the light panel either remained red or gradually shifted to green except for continuous minor fluctuations that coincided with real S lever movements. Each discussion lasted between 12 and 14 minutes. During each discussion the *T* brought out four reasons for preferring radio: (1) radio has greater coverage in our state and would reach a larger audience; (2) more Ss could participate because radio time costs less and there would be more contests; (3) radio shows will take fewer rehearsals and impose on their free time less; and (4) a participant will get nervous on TV with thousands watching every move; on radio their team could give a better performance.

6. After the discussion a paper-and-pencil 'opinionaire' was filled out by the Ss. This paper-and-pencil instrument permitted Ss to (a) indicate their preference for radio vs. TV; (b) indicate their awareness of the movement of the light panel; (c) give an indication of their interest in further participation; (d) react to three items to test their cohesiveness; (e) give an indication of their dependency on the teacher; and (f) indicate those members of their group that they thought would do the best job on the program (sociometric). As soon as the students had filled out the post-discussion questionnaire the *T* collected them and left the room as the *E* re-entered.

7. The *E* told the Ss that the sponsor wanted to make sure that each team had the best possible opportunity to compete. Therefore, if they had any reservations at all about working with this particular *T*, it would be possible to make a change. There were quite a number of *T*s employed for this purpose and this particular one need not necessarily be assigned to their group. In order to find out their reactions they were asked five questions. These items included: (a) whether they would like to continue to work with the same *T*, (b) how well he explained the radio vs. television issue, (c) whether or not he was fair, (d) whether he appeared to be more interested in each individual or mostly interested in the group as a whole, and (e) whether they were aware of the fact that the teacher-trainer preferred either radio or television.

8. The *E* then went through a carefully planned deflation procedure that revealed the deceptions used in the experiment and also notified the Ss that there would be no competition. The nature of the experiment and the reasons for using deception were explained. During this 'decompression' session each S was encouraged to express his feelings about the experiment. Each S was requested to cooperate in not revealing the nature of the experiment to

others who might become Ss, and there were usually verbal pledges given for cooperation in this respect. The students were then taken back to their school.

## The Analysis of the Data

### The Experimental Population

The experimental population consisted of 360 students from seven different high schools in Minneapolis and St. Paul, Minnesota. Local high-school principals or counselors selected groups of ten such that (a) within any one group friendships and acquaintanceships would be at a minimum, and (b) among groups academic ability would be randomly distributed. This necessitated contacting large high schools that had ten or more tenth-grade home-rooms in order to form groups that approached the two specifications by selecting one *S* from each home-room. Three of the high schools were in above average socio-economic areas, two were definitely below average, and two were average. The average age of the Ss was about sixteen years. There are no reasons to suspect that the experimental population differed significantly from a random sample of urban high-school tenth-graders.

### Evidence of T Consistency Between Treatments

The design called for an individual approach by the *T* in treatments I-O and I-C, a group approach in treatments G-O and G-C, with other aspects of the *T*'s behavior to be as consistent as possible. Seven trial runs with different Ss were used to train the *T* to behave consistently in each approach; the same trained *T* was used in each of the 36 discussions to control personality factors. During each discussion the *E* used electrical recording equipment to keep an accurate record of the total time the *T* talked as well as the Ss. Later analysis of the records showed that no statistically significant variations of the total time the teacher talked occurred between treatments. In addition, the *T*'s verbal statements were classified according to whether they were directed toward individuals or toward the group as a whole. The results showed that the *T*'s approach varied in a manner consistent with the experimental design.

As a further check on the consistency of the *T*'s behavior, all the Ss rated the *T* immediately after he left the room, using the item, 'The teacher-trainer helped explain the advantages and disadvantages of radio and television so that I: (1) understood it very well, (2) understood it fairly well, (3) am not sure that I understand it, (4) am puzzled about a number of points, or (5) do not understand it at all.' Assigning values of one through five to the responses, the mean scores for the various treatments were calculated, and no statistically significant differences were found between treatments.

All data concerning the consistency of the *T*'s behavior are not presented here, to conserve space. An analysis of these data is available upon request. The authors conclude that there were no known behavior differences of the *T*, beyond those required by the individual and group approaches, that could reasonably bias the variables under study.

## The Effectiveness of the Experimental Manipulations

In an experiment that is complicated by the attempt to create a special social situation complete with particular individual perceptions, it is reasonable that some manipulations would fail to convince some Ss. A number of Ss did not respond as planned, and thus they existed in a situation that is not a proper test of the hypotheses of this study.

First, three groups were discarded because: one group was highly suspicious, nervous, and obviously upset—this was later traced to an unfortunate statement made by a counselor just as the Ss left their high school; a second group, from a low socio-economic district, contained two uncooperative Ss who refused to follow directions, used profanity, and tried to smoke during the experiment; and a third group was discarded because it showed unmistakable signs of knowing about the deceptions of the experiment prior to participating. The elimination of these 30 Ss from the study reduced the population to a total of 330 Ss.

Second, a total of 87 Ss indicated a preference for radio before the discussions took place. They obviously could not be influenced to shift their preference from TV to radio. The incidence of these Ss by treatment is shown in column 2 of Table 1.

TABLE 1. Incidence of Misperceivers by Treatment

| Treatment (1) | Preferred Radio (2) | Other Misperceptions (3) | Potential Shifters (4) | Total (5) |
|---|---|---|---|---|
| G-O | 14 | 16 | 50 | 80 |
| I-O | 23 | 15 | 52 | 90 |
| G-C | 23 | 28 | 29 | 80 |
| I-C | 27 | 15 | 38 | 80 |
| Total | 87 | 74 | 169 | 330 |

$\chi^2 = 18.36$; at 6 d.f., $p < 0{\cdot}01$

Third, a total of 74 Ss failed to respond properly to some other manipulation: 28 failed to see the *T*'s preference for radio; 41 failed to show understanding of the light panel; and 5 failed at both perceptions. The incidence of these misperceivers by treatment is shown in column 3, Table 1.

The net result of the manipulations is shown in column 4 of Table 1, which indicates that 169 Ss were in the social situation described at the beginning of this article. The analysis of compliance to test our major hypotheses will make use of these 169 cases.

It would be nice if these misperceptions had occurred without any relationship to the experimental treatment. A simple contingency calculation of Chi-square, testing whether there is any particular relationship among the cells of Table 1, yields a Chi-square of 18·36, which is significant, at six degrees of freedom, beyond the 0·01 level. The incidence of misperceivers is higher whenever the light panel is used to indicate a shift of group opinion toward radio (Treatments G-C and I-C) and is particularly high in the G-C Treatment. No plans were made to collect data that might reveal why Ss misperceived the manipulations; therefore the significant differences revealed in Table 1 cannot be interpreted systematically. The most obvious explanation is that most Ss held their preference for TV strongly and were not willing to accept evidence of group opinion shift toward a preference for radio.

### Compliance in the Different Treatments

Compliance was measured as a shift in opinion from TV to radio on the pre- and post-discussion ballots. Hypothesis 1.0, predicting more compliance in the individual treatments I-O and I-C, can be tested by analyzing the opinion shift of the 169 Ss shown in Table 2. Hypothesis 2.0, predicting more compliance in the compliance treatments G-C and I-C, can be similarly tested. First, a 2 × 4 Chi-square calculation of the entire table rejects the hypotheses that there are no differences among the cells, $p<0\cdot0001$. A 2 × 2 Chi-square test of the greater compliance in the I-O and I-C Treatments rejects the hypothesis of chance differences, $p = 0\cdot018$, thus supporting hypothesis 1.0 regarding teacher influence. Finally, a 2 × 2 Chi-square test of the greater compliance in Treatments G-C and I-C also rejects the hypothesis of chance difference, $p<0\cdot001$, supporting hypothesis 2.0 concerning group contagion.

TABLE 2. Opinion Shift of Subjects Who Perceived Manipulations, Prediscussion Preference Compared to Post-Discussion ($N = 169$)

| Treatment | Preferred TV Before Discussion Potential Shifters (1) | Shift to Radio (2) | Per Cent Shifting (3) | Seat Position Odd (4) | Even (5) |
|---|---|---|---|---|---|
| G-O | 50 | 7 | 14·0 | 3 | 4 |
| I-O | 52 | 10 | 19·2 | 3 | 7 |
| G-C | 29 | 14 | 48·3 | 6 | 8 |
| I-C | 38 | 22 | 57·9 | 13 | 9 |
| Total | 169 | 53 | | 25 | 28 |

$\chi^2=26\cdot6$; at 3 d.f., $p<\cdot001$

Limiting the test of our major hypotheses to the 169 Ss in Table 2 may be questioned by some readers. The data for all 330 Ss are not shown, in order to conserve space, but the trends and differences for all 330 Ss are similar to those for 169 Ss. The per cent shifting opinion, for example, for 330 Ss are: G-O, 19·7 per cent; I-O, 23·9 per cent; G-C, 29·8 per cent; and I-C, 43·4 per cent. The same Chi-square tests for the 330 Ss are: for the entire 2 × 4 table, $p = 0 \cdot 03$; for teacher influence, $p = 0 \cdot 15$ (not significant); and for group contagion, $p = 0 \cdot 114$. Apparently, when Ss who either did not understand or refused to accept the experimental manipulations are added to those Ss who did, the trends remain the same, and only the differences regarding teacher influence become non-significant.

These data on compliance show that when a *T* talks to the Ss 'one at a time' and, at the same time, the individual Ss believe that the rest of the group is complying to the *T*'s influence, then maximum compliance occurs. When the *T* talks to the group and most members of the group believe that others are not complying to the *T*'s influence, then minimum compliance occurs. Finally, when the force of the *T*'s influence is compared with the forces of group contagion, the latter are more effective in causing compliance.

### The Effect of the T's Behavior on Cohesiveness

Hypothesis 1.1 predicts the cohesiveness will be lower in the individual treatments. Two post-discussion questions used to measure cohesiveness reflect the desire of the Ss to remain in their group. (1) 'If the group should decide to broadcast over TV or radio in a way that is the opposite to what you prefer, would you like to remain with this group or transfer to another group from your school?' (2) 'Suppose you do not get your choice about being a participant or a helper. In such a case would you like to remain with this group or transfer to another group from your school?' The response form permitted a S to indicate his choice.

TABLE 3. Subjects Who Desire to Transfer out of Group

| Treatment | N (1) | Choices to Transfer Odd (2) | Even (3) | Total (4) | % Inds. in Each Treatment* N=65 (5) |
|---|---|---|---|---|---|
| G-O | 80 | 2 | 3 | 5 | 7·7 |
| I-O | 90 | 15 | 15 | 30 | 46·2 |
| G-C | 80 | 10 | 7 | 17 | 26·1 |
| I-C | 80 | 5 | 8 | 13 | 20·0 |
| Total | 330 | 32 | 33 | 65 | 100·0 |

* The percentage distribution of column (4).

The results from the two questions on cohesiveness are presented in Table 3 and show that 65 Ss among the 330 indicated a desire to transfer from the group. When Treatments I-O plus I-C (individual approach) are compared with G-O plus G-C (group approach), there are about twice as many respectively who wish to transfer. The difference (43 compared to 22) is significant beyond the 0·01 level of confidence and supports hypothesis 1.1. As expected, there are no significant differences when this same comparison is made between Treatments G-O plus I-O vs. G-C plus I-C.

## Evidence of Dependency on the Teacher

Hypothesis 1.2 predicts that the individual approach will create more dependency than the group approach. One aspect of dependency is measured by the post-discussion question, 'Would you like the teacher or the group to select the five "program participants"?' The percentage of Ss ($N = 330$) requesting that the *T* make this choice in each experimental treatment is: I-O, 40 per cent; I-C, 37·8 per cent; G-O 32·5 per cent; and G-C, 31·3 per cent. This measure of dependency is higher for the individual treatments, but the difference could have occurred by chance about one out of three times, $p = 0 \cdot 3$. The individual or group approach, apparently, did not significantly influence the responses of the Ss to this question.

Additional evidence on dependency can be secured, however. One test can be made within the individual treatments, ignoring the group treatments. One can reason that the *T*'s individual contacts with Ss in the odd-numbered seats should make these Ss more dependent that Ss seated in the even-numbered seats. During each contact the *T* would try to praise the individual, yet at the same time turn his ideas into arguments for radio; failing this, the *T* would confront the S with counterarguments in favor of radio. During these contacts the *T*'s attention was centered on the individual. The Ss in the even-numbered seats were only spectators to these more personal experiences. Data from the same question (see above) are shown in Table 4 for all Ss in the individual treatments. A significantly greater ($p<0 \cdot 01$) number of Ss seated in the odd-numbered seats wanted the *T* to select the five program participants, which tends to support hypothesis 1.2.

TABLE 4. Preferences for Appointing Program Participants, Odd Versus Even Seats in Individual Treatments

| | Choice | | |
|---|---|---|---|
| Seat | Teacher | Group | Total |
| Odd | 40 | 42 | 82 |
| Even | 25 | 62 | 87 |
| Total | 65 | 104 | 169[*] |

[*] One student did not answer. Total *N* should be 170; 90 Ss in treatment II; 80 Ss in treatment III.

Finally, it is shown in another paper growing out of this study that those Ss who sat in the odd-numbered seats and talked with the teacher were chosen significantly ($p = 0{\cdot}01$) more frequently than others in response to the question, 'which five Ss would make good program participants?' Since these are the same Ss who, in turn, gave evidence of dependency on the *T*, the overall relationship seems to be: the *T* praises and argues with certain Ss; these same Ss are more likely to show dependency on the *T*; these same Ss are chosen by others for positions of responsibility.

### Spontaneous Lever Movements

One can accept the divisive effects of the *T*'s influence on cohesiveness, the development of dependence and compliance, and still wonder about the more private, inner reactions of the Ss. In particular, we wish to turn to the Ss who shifted their opinions from TV to radio. Did these opinion shifts occur with private, inner acceptance or with private inner resistance? 'Inner acceptance,' in this discussion, is defined as not only being satisfied with one's shift of opinion, but, in addition, actively influencing others in the group to shift to radio. 'Inner resistance,' similarly, means not only being dissatisfied with the shift to radio, but, in addition, actively influencing others in the group to resist this change. The spontaneous lever movements that were recorded when the Ss thought they were secretly influencing the group's opinion, provide considerable evidence on this question.

Interviews and post-experimental discussions indicate that none of the Ss realized that their lever movements were being recorded. The lever movements could be made secretly and the Ss could not trace lever movements, even their own, by observing the light panel. Their 'spontaneity' resulted primarily from the Ss' perception that they could use the lever to influence the immediate situation.

Before turning to the data, a word about their form. Lever movements were averaged for three-minute periods. Average lever movements were calculated by adding all lever movements for radio and dividing by 10, the number of Ss in a group for each three-minute period. TV average lever movements were calculated in an identical manner. The standard errors of the means vary from 0·2 as high as 2·5 depending on the number of Ss involved. For $N = 15$ or larger the standard error is usually less than 0·5. An extensive statistical treatment of these averages does not seem justified in terms of the limited inferences to be made. Simple *t*-tests of the significance of differences between means have been used. Whenever the word 'significant' is used in reference to the comparison of lever movements, the *t* calculation indicates a *p* value of 0·01 or less.

The first question that can be asked is, 'Are the lever movements valid?' Any question about the validity of lever movements really refers to what lever movements are supposed to express compared with what they do express. According to the directions given to the Ss, lever movements should be an expression of personal opinion. One way to check this is to compare the lever movements of Ss who voted consistently in favor of radio with those of Ss who consistently voted in favor of TV. We would expect to find predominately TV lever movements for pro-TV Ss and predominately radio lever movements for pro-radio Ss. Average lever movements per three-minute period for Ss who voted TV-TV are shown in Figures 1 and 2; those for Ss who voted radio-radio in Figures 3 and 4. The data are grouped according to the opposition and compliance conditions, because the differences attributed to the *T*'s approach are insignificant.

When Figure 1 is compared with Figure 3, and when Figure 2 is compared with Figure 4, the TV-TV Ss have significantly more TV lever movements and the radio-radio Ss have significantly more radio lever movements. It is clear from these comparisons that Ss used the levers to express their preference for either TV or radio.

Another aspect of validity is the question, 'Did the Ss attempt to influence others in the group?' Evidence concerning this question can also be obtained from Figures 1 through 4. First, we would predict that Ss who preferred TV would express more lever movements in the compliance condition when it appeared that the group was shifting to radio; in the opposition treatments they could tell that the group still preferred TV and would be less inclined to use their levers to influence the group. The fact that there are significantly more TV lever movements in Figure 2, compared with Figure 1, suggests that the Ss did try to influence the group by using the levers. For those Ss who preferred radio, we would expect more radio lever movements during the opposition condition when it appeared that the group was not shifting to radio. There are slightly more radio lever movements in Figure 3 compared with Figure 4, but this difference is not significant. Perhaps the awareness that such efforts were not just resisting the group, but in addition an aid to the *T*'s arguments, may have curtailed the lever movements of pro-radio Ss. Nevertheless, we can assume that those Ss who prefer TV can and do use their levers in an attempt to influence the group.

The total picture from Figures 1 through 4 strongly supports the validity of average lever movements. To these objective comparisons could be added the subjective impressions of the *E* who observed all sessions and whose own light panels, in the observation room, indicated the true lever movement of the Ss. To him the spontaneous lever movements seemed both logical and persuasive. Ss with strong convictions 'rode' their levers in an effort to influence the group. The authors have accepted spontaneous lever movements as valid indications of the secret preference of the Ss and their efforts actively to influence others.

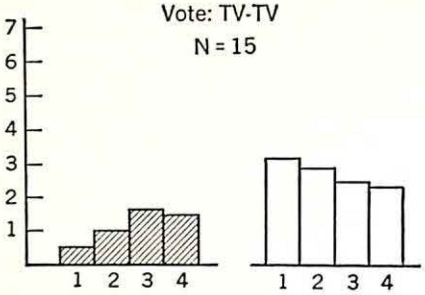


Figure 1
Opposition Treatments

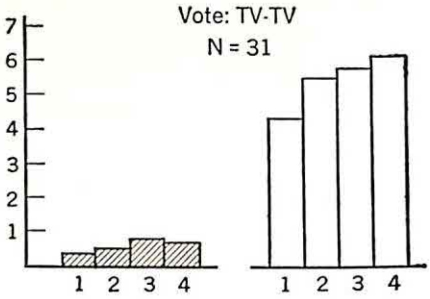


Figure 2
Compliance Treatments

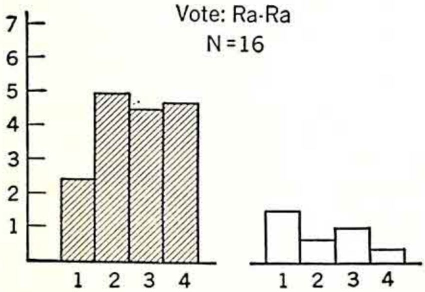


Figure 3
Opposition Treatments

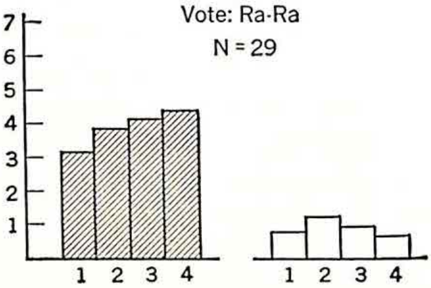


Figure 4
Compliance Treatments

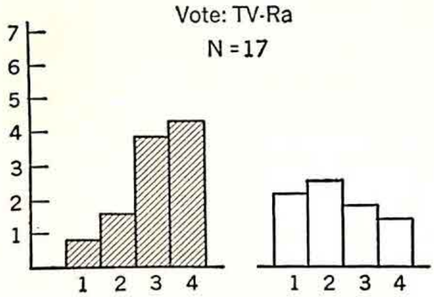


Figure 5
Opposition Treatments

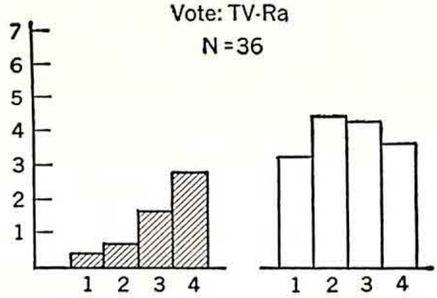


Figure 6
Compliance Treatments

FIG. 1–6. *The four time periods of three minutes each are indicated by the use of a separate bar. The average of radio lever movements during the first period is shown by the first shaded bar on the left, the average of TV lever movements during the same period is shown by the first clear bar on the right. The pre- and postdiscussion voting pattern is indicated on each figure; e.g. Figure 2, 31 Ss voted for TV before and after the discussion in the compliance treatments; during the second time-period they averaged about one-half lever movement per person for radio and, during the same period, the same Ss averaged about 5·6 lever movements for TV.*

Next we turn to the average lever movements of Ss who shifted their opinions from TV to radio and thus complied with the influence of the *T*. Their average lever movements are shown in Figures 5 and 6. In both these figures, there is an orderly increase in the number of lever movements in favor of radio as the discussion proceeds. In all treatments the highest number of radio movements occurs in the last three-minute period. This must be due to the logical persuasion of the *T*'s arguments. For the 17 Ss in the opposition condition, Figure 5, their shift of opinion seems to occur with inner acceptance. This inference is made from the fact that the TV lever movements decrease as the radio lever movements increase, so that in the final three minutes of discussion, their radio lever movements occur three times as frequently as their TV lever movements. These Ss shift their opinions in spite of the apparent opposition of the group and are not conforming to group pressures but are complying to *T* persuasion.

In Figure 6, when the group was apparently complying to the *T*'s influence, the proportion of TV and radio average lever movements is significantly different. At all times the average lever movements for TV exceed those for radio. The consistent inference here is that these Ss wrote down their final vote for radio with considerable inner resistance and perhaps even a secret preference for TV.

Combining the TV average lever movements for the G-C and I-C Treatments in Figure 6 does hide one significant difference between these two treatments. The TV average lever movement in the I-C Treatment is 4·17, the same average for the G-C Treatment is 3·12. This difference is significant and means that the highest inner resistance of all occurs in the I-C Treatment. In both of these treatments Ss who shift are conforming to apparent group opinion as well as complying to *T* persuasion.

One final piece of evidence about active resistance can be obtained by making a graph that shows the average lever movements separately for Ss seated in the odd- and even-numbered seats. This was done by excluding Ss who voted for radio both before and after the discussion, since their TV lever movements are very infrequent, as is shown in Figures 3 and 4. Next, the graphs for all Ss participating in the treatments involving the group approach of the *T* are not shown because no significant differences appear or would be expected. There remains 52 Ss in the I-O Treatment whose radio and TV average lever movements are shown in Figure 7, and 38 Ss in the I-C Treatment whose same lever movements are shown in Figure 8.

The differences between the odd and even seats are not significant for radio and TV average lever movements in Figure 7 and also for the radio movements in Figure 8. However, the TV lever movements in Figure 8 averaged over all four discussion periods are 4·71 for Ss in the odd-numbered seats, which is significantly different from the average of 3·75 for Ss seated in the even-numbered seats. Inner resistance is not only highest in the I-C Treatment, as reported earlier, but it is also highest for Ss seated in the odd-numbered seats who participated in the verbal interaction.

To summarize, the lever data shed light on the inner acceptance or rejection of the pressures of group conformity as well as teacher compliance. It is clear that those Ss whose pre- and post-opinions are consistent, voting either TV-TV or Ra-Ra, expressed stable, inner convictions in their use of the levers. These Ss, who were in the majority, resisted either the *T* or the group according to their own preferences and the treatment.

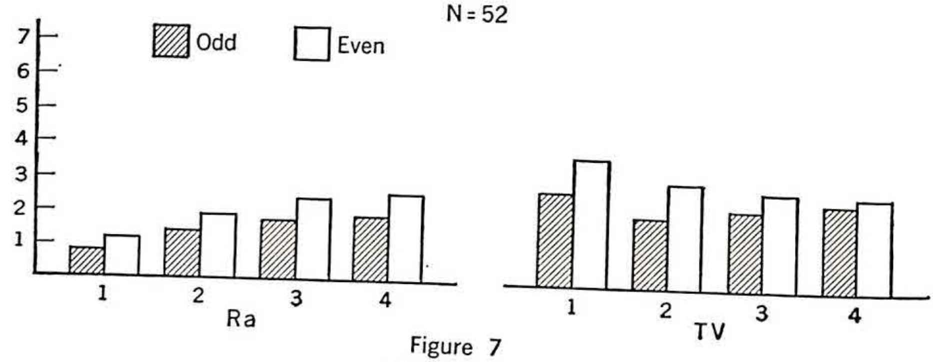


Figure 7
I-O Treatments

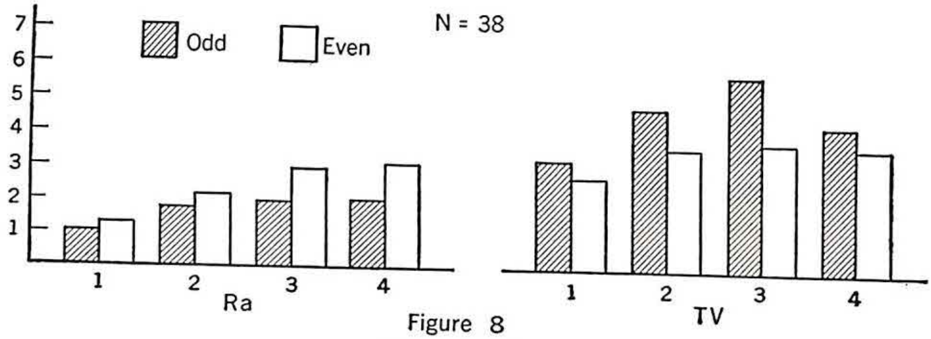


Figure 8
I-C Treatments

FIGS. 7 and 8. *The average of lever movements for Ss in the odd and even seats are shown separately in these two figures. For each three-minute time-period separate bars show the odd-even averages for radio, on the left, and for TV, on the right.*

Those few Ss who shifted from TV to radio in the opposition treatments in spite of perceived group resistance, appear to be logically persuaded and are willing to ignore or reject group pressure to follow the teacher. In the case of the compliance treatments, when the group appeared to be shifting, more Ss also shifted publicly, but inner resistance to their own compliance and conformity is clearly visible. The *Ss* seated in the odd-numbered seats, in direct verbal contact with the *T*, show the highest inner resistance of all.

## Conclusions and Interpretations

### Conclusions

Hypotheses 1.0 and 2.0, concerning the effects of the *T*'s influence and group contagion on compliance, are strongly supported. Equally clear evidence can be found to support hypothesis 1.1 that the individual approach of the *T* is divisive and reduces group cohesiveness. Fairly strong evidence has been found to support hypothesis 1.2 about the development of dependency as an outcome of restrictive *T* domination.

The logical persuasion of the *T* appears as a constant force toward compliance in all four treatments. With the group approach, this force is reduced somewhat by an implied support of the group caused by the *T*'s group approach. The other major social force affecting opinion shift is perceived group opinion. In the opposition treatments, this force favors resisting compliance; in the compliance treatments this force favors compliance.

Under conditions of this experiment, the more potent of these two forces is the force of group conformity. It is not surprising that *Ss*, brought together as a team to represent their school, are more sensitive to their group opinion than they are to the efforts of a strange *T* who presents four logical arguments in the short period of 12 or 13 minutes.

In Treatment G-O, 85 per cent ($N = 50$) of the *Ss* do not change their preference for TV. The freedom to express ideas spontaneously appears to lend support to the group, in spite of the *T*'s arguments. In this treatment the fewest number of *Ss* wish to transfer to another group and dependency on the *T* is lowest. Apparently dependency is less likely to develop in the presence of strong group resistance. The few *Ss* who do shift their opinion seem to be logically persuaded with inner conviction.

In Treatment I-O, 80 per cent ($N = 52$) of the *Ss* keep their preference for TV. Freedom to express ideas is restricted to those in the odd-numbered seats and then only with the permission of the *T*. The *T*'s persuasive efforts center on individuals and ignore the group as a whole while the group appears to oppose shifting to radio. This treatment has the highest incidence of *Ss* desiring to leave the group (see Table 3). Perhaps this is due to an increasing anxiety, which some *Ss* might feel, because the more potent individual approach is being openly resisted by *Ss*.

In Treatment G-C, 48 per cent ($N = 29$) shift their opinion from TV. The greater impact of group conformity is clear, for those who shifted publicly. But for those who preferred TV initially, including those who shifted, private resistance increases, judging by the higher TV lever movements compared with the first two treatments. For the Ss who shifted, this can be interpreted as a lack of inner conviction, almost ambivalence. For those who did not shift from TV, this can be interpreted as resistance to group conformity as well as resistance to *T* compliance.

In Treatment I-C, 58 per cent ($N = 38$) shift from TV preferences. Here the 'divide and conquer' mechanism seems to operate. The *T*'s restrictive, persuasive methods appear to be supported by a total group shift of opinion, as indicated by the light panel. Increased dependence on the *T* appears as a trend. For those Ss who did shift, lack of inner conviction is the highest of all treatments. In fact, during the last three-minute period, those who publicly shifted in favor of radio registered more lever movements for TV compared with radio. *Apparently, the situation that produced the highest public compliance also produced the greatest inner resistance for those shifting.*

## Implications

This experiment was conducted in the context of teacher-pupil relationships. Our findings contain some implications for classroom situations in which a similar issue is involved.

1. A teacher who sets out to persuade a class to shift its opinion by 'divide and conquer' technique is likely to find that this method is most successful, but, at the same time, very likely to produce inner resistance. It is reasonable to expect that such resistance would be expressed sooner or later if the teacher remained in contact with the class over longer periods of time than obtained in this experiment.

2. In situations where students can judge and discuss the effectiveness of the teacher's persuasive efforts, the success of the persuasion will be influenced by the freedom with which students share their perception of the teacher's efforts. A perception of teacher ineffectiveness is likely to create group pressures strong enough to resist the logical persuasion of the teacher, even in groups that are newly formed. It seems reasonable to expect that these group pressures would be greater in classrooms where the students had the opportunity to develop stronger bonds of friendship and class cohesiveness.

3. By implication, this study clarifies a kind of vicious circle that exists in some classrooms in which teachers assume a dominant, friendly, and successfully persuasive role centered primarily on individuals. In this role the teacher is most likely to achieve temporary success and thus feel rewarded. But when the inner resistance develops and is expressed, the teacher is likely to reassert his authority in the same fashion and this tends to sustain class resistance. On the other hand, when the goals of the teacher and the class are not in conflict and the teacher's efforts are supported by the development of group pressure, then the strongest influence forces are created.

# 18

# Use of Verbal Reinforcement in Developing Group Discussion Skills — *M. Ray Loree and Margaret B. Koch*

*Prerequisite to the attainment of "complex educational objectives," using principles of reinforcement, is the identification of adequate responses in movement toward those objectives. Besides this, however, are other important questions surrounding the application of principles derived from reinforcement learning theories. The present authors have investigated one approach to answering these questions as they relate to the development of group discussion skills. Since in their experiment immediate reinforcement could not be given (it would have involved interruptions), the authors used tape recordings of the group discussions. Later, when playing them back for the experimental group, the instructor would stop the recordings whenever students made "adequate" responses, in order to verbally reinforce them. Another group was instructed in the characteristics of good discussions and in the scoring criteria, but did not hear playbacks. The experimenters found in the final discussions significant learning differences between the two groups that favored the experimental one. Some interesting implications for teacher behavior in situations involving classroom interactions are suggested by this research.*

THE APPLICATION of principles derived from reinforcement learning theories to the task of guiding students' attainment in the more com-

Reprinted from *Journal of Educational Psychology*, 1960, **51**, 164–168, by permission of the senior author and the American Psychological Association.

plex educational objectives of the school presents many problems. How does the social science teacher who wishes to develop the students' ability to think more effectively on social science issues go about identifying the responses or the classes of responses to be reinforced? What are the kinds of appropriate and effective rewards that can be manipulated by the teacher in the classroom? Is it possible for the composition teacher to provide immediate reinforcements for the host of responses made by the student in the process of writing a theme?

Certain research findings suggest possible avenues of investigation that may shed light on the above questions. The technique employed by Bloom and Broder (1950) constitutes one approach to the problem of identifying adequate response patterns in complex verbal problem solving. Responses of "good" and "poor" problem solvers as they "thought aloud" while working on examination questions were compared and analyzed, and crucial differences in their problem solving processes were identified. Some technique designed to identify adequate responses would seem to be a first prerequisite to any effort to utilize reinforcement concepts in guiding problem solving type learning. For certain kinds of complex learning a second prerequisite may be the identification of what Skinner (1953) refers to as the "contingencies of reinforcement"—the responses that need to be reinforced in order to increase the probability of the occurrence of the desired response pattern.

What rewards may be manipulated by the teacher in order to reinforce appropriate responses? Obviously the rewards most frequently used in simple learning experiments, i.e., food and water, have little if any appropriateness for the classroom. However, it is well established that another form of reward—"knowledge of results"—facilitates learning (Wolfe, 1951). Studies on the conditioning of verbal behavior indicate that simple verbal cues such as "good" or "mmm-hmm" may serve to reinforce specific or even generalized response patterns, and that this reinforcement occurs even when the subjects of the experiment are not aware that certain of their responses are being reinforced (Krasner, 1958; Salzinger, 1959). Additional possibilities occur when students are working in groups. Because individuals differ in personality, personal goals, and in whether they are task oriented, self-oriented, or interaction oriented (Fouriezos, Hutt, & Guetzkow, 1950), it may be expected that in a group problem solving situation, what constitutes reinforcement will vary from individual to individual.

The relative merits for school learning of immediate as opposed to delayed reinforcement have not been fully explored, although immediate reinforcement is usually assumed to be the more effective. However, it is difficult to implement immediate reinforcement for some kinds of school learning. The "teaching machines" described by Skinner (1958) represent one promising approach. The "stimulated recall" technique described by Bloom (1953) may be a means through which immediate reinforcement conditions may be simulated. Bloom reports that students, listening to a play-back of discussion and lecture sessions in which they participated, seem to relive the experience. Utilizing the "stimulated recall" technique in a group problem solving session, it should be possible to simulate immediate reinforcement conditions by interrupting the playback and providing verbal cues whenever a member of the group exemplifies good problem solving behavior.

## Purpose

The purpose of the present study is to investigate the effectiveness of positive reinforcement in developing certain group discussion competencies of students. An effort was made to simulate "immediate" reinforcement through the use of the stimulated recall technique. Topics such as: "What are the possibilities of raising a child's I.Q.?" "How can we prevent forgetting?" "Of what psychological significance are the body changes that take place during adolescence?" were used for group discussion. References for each topic were provided. Group discussion competencies were assessed on the basis of performance on a 10-min. panel discussion.

## Subjects

Ninety-six undergraduate students enrolled in four educational psychology classes at Louisiana State University constituted the Ss in this study.

## Procedure

Each of the four classes was divided into six panel discussion groups of four members each. Students were assigned to discussion groups in such a way that the average academic grade point average for each group was approximately equal (the range in the group mean grade-point average for the 24 groups was 1.5 to 1.6). One student with a very good academic record was assigned to each group and was designated as a group leader.

The experimental instructional conditions for each class were assigned thus: the two classes of one instructor were designated as the *control non-practice groups*; on the basis of a coin toss, one class of a second instructor was designated as the *experimental reinforced groups* while the second class of this instructor was designated as the *experimental practice groups.*

Six pairs of discussion topics were selected for use in initial and final discussion sessions. For each topic assigned, a number of references were listed.[1] An effort was made to equate type of content and difficulty level for each pair of topics. For example, Topic 1 was: "Are newer methods of teaching reading superior to older methods?" while the paired Topic 2 was: "What is the evidence concerning the effectiveness of newer methods of teaching arithmetic?" One experimental reinforced group (A) and one control group (B) were assigned Topic 1 for their initial panel and Topic 2 for their final panel. An experimental practice group (C) and one control group (D) were assigned Topic 2 initially and Topic 1 for their final panels.

For the initial panel discussion sessions, the same procedure was followed in each of the four classes. After the instructor explained in each class that an opportunity would be provided in the course for students to develop their skills in applying psychology to educational problems, the panel discussion procedure was explained and each student was assigned to a discussion group. Topics were assigned and each group met for a few minutes to divide up the suggested references. Students were requested to have their readings on the topic completed for a class session a week hence, when they would be given a class period to plan for a 10-min. panel discussion that was to be tape recorded. The whole procedure was presented in terms of an instructional technique to reach important educational objectives of the course rather than in terms of a research study. Two days after the class period devoted to planning, groups were scheduled to meet in offices and carry out their 10-min. panel discussions. At this point in the experiment the procedure was varied for each of the four classes.

For the control nonpractice classes no further panel discussions were assigned until the final panel toward the end of the semester. It was thought that possibly a course in educational psychology might contribute to skillful performance in panel discussions on topics in educational psychology. For the two experimental classes two additional panel discussion topics were assigned to each group during the semester. The topics were of a similar nature to those used in the initial and final panel discussions. The method of assigning and carrying out the additional panel discussions was identical for both experimental classes. The difference in treatment was in having each group in the reinforced class listen to a playback of its panel. All four classes were informed that their performance in the final panel would be graded and counted as part of their final course grade.

In the experimental practice class, after each of the panel discussion sessions, the instructor talked only in general terms about the performance of the groups on the panel. However, quite specific information was given to the class on the characteristics of a good panel discussion. The categories, later used in rating transcripts of initial and final panels, were discussed. It was explained that a good panel discussion contained an introduction, a main body, and a conclusion. Each of these aspects of a panel discussion was discussed in detail. It was explained, for example, that an introduction serves the purpose of making clear the kind of information needed to discuss the topic; that sometimes there was need to define key terms of the topic; and sometimes it was advisable to delimit the topic. No playback of the panel was provided, however, and specific knowledge of results was not provided to each group.

---

[1] The topics used along with their references may be found under "Discussion Questions" in M. R. Loree (Ed.), *Educational psychology*, New York: Ronald Press, 1959. Topic number, question number, and page reference for each of the twelve topics are as follows: 1 - 2 - 30; 2 - 3 - 30; 3 - 1 - 101; 4 - 2 - 101; 5 - 1 - 138; 6 - 2 - 138; 7 - 2 - 275; 8 - 5 - 275; 9 - 1 - 328; 10 - 4 - 328; 11 - 2 - 328; 12 - 3 - 328.

The instructor in the experimental reinforced class scheduled each of the groups to meet in an office to hear a playback of the panel discussion. During the playback, the instructor stopped the recording at key points—whenever a panelist was caught employing good panel discussion techniques. The instructor would then make comments such as:

> Good, you have defined a key term in this topic. . . .
>
> Good, your report of this research study is succinct, yet you have brought out the information that is relevant. . . .
>
> Notice how the question asked by the last speaker leads to a clarification of the point previously made.

Only positive reinforcement was provided, although after the members of the group had listened to the complete playback an evaluation of the panel as a whole was given and weaknesses were pointed out.

### Treatment of Data

Transcripts of the 24 initial and the 24 final discussion sessions constitute the raw data for this study. These transcripts were coded so that raters could not tell whether the discussion represented an initial or a final panel, a panel of a reinforced group, a practice group, or a control group.

A panel discussion rating form was devised that included the following categories:

Part I—Introduction

Part IIA—Main Body ratings for individual panelists, including ratings on:

1. Extracting relevant material from references
2. Reporting accurately contents of reference
3. Avoiding irrelevant details
4. Identifying source of evidence
5. Utilizing experimental evidence rather than personal opinion

Part IIB—Main Body ratings on the panel as a whole, including ratings on:

1. Interpreting the reference material
2. Organizing materials

3. Interacting rather than merely presenting a series of speeches

Part III—Summary, including ratings on:

1. Identifying essential elements of the problem
2. Referring to the major findings contained in the main body
3. Evaluating the completeness of the discussion.

Each of the 12 categories was rated on a 13 point scale except the category "identifying source." This category was rated on a three point scale. The individual Main Body ratings were averaged for each panel. Thus, it was theoretically possible for a panel to obtain a total rating as high as 146 points.

Prior to the actual rating of the transcripts, two graduate students were each given six of the discussion topics and asked to extract for each listed reference the material relevant to the discussion topics. The principal investigators then took these summaries and defined as explicitly as possible a scoring system for each category. For the relevancy score, for example, six major points might be extracted from a particular reference. The numerical ratings to be assigned for all possible combinations of the six major points were then agreed upon by the two principal investigators. The task of the two raters became then to independently rate each panel utilizing the agreed upon rating plan. The total time consumed by the two principal investigators in rating was about 200 hours. A Pearson product-moment correlation coefficient of .972 was obtained for the independent ratings. After the four transcripts for one topic were rated, the two raters went over their ratings together and, where there were discrepancies, came to an agreement rating.

## Results

Although groups were equated for academic ability, there was no way to equate groups for their initial ability to perform in a panel discussion. An analysis of variance of the four groups indicates that, although differences on the initial panel discussions exist in the means of the groups, these differences are not significant ($F = .00$). However, to adjust the final scores for the small differences existing, an analysis of covariance was used (Edwards, 1950). This analysis indicates that the difference between groups is significant ($F = 15.90$, $df = 3$ and $19$, $p < .01$). Table 1 shows the initial, final, and adjusted means for the six panels in each of the four classes.

## Summary

This study is concerned with the effect of positive reinforcement in developing certain group discussion competencies.

TABLE 1. Initial and Final Means and Standard Deviations of Reinforced, Practice, and Control Classes

| | Reinforced Groups (A) | Practice Groups (C) | Control Groups (B) | Control Groups (D) |
|---|---|---|---|---|
| Initial Means | 70.5 | 66.0 | 65.0 | 70.2 |
| *SD* | 18.6 | 6.7 | 11.7 | 7.1 |
| Final Means | 103.5 | 68.5 | 67.7 | 70.2 |
| *SD* | 12.6 | 12.0 | 8.4 | 8.0 |
| Adjusted Means | 102.4 | 69.3 | 68.9 | 69.3 |

Each of four educational psychology classes was divided into six panel discussion groups. Each group conducted a panel discussion at the beginning of the semester and, on a matched topic, at the end of the semester. All panel discussions were tape recorded, and later transcribed, assigned a code number to insure anonymity, and rated independently by two judges. An agreement rating was obtained where discrepancies in judgments between raters occurred. The agreement ratings on the final panel discussions constitute the dependent variable for this study. The independent variable is the type of training provided between initial and final panel discussions. No additional practice in participating in panel discussions was provided for the groups in the two "control" classes. Two additional panel discussions were conducted by the groups in the two experimental classes. For the "practice" groups the discussions were tape recorded but not played back to the students nor evaluated as to their quality. Discussion of the "reinforced" groups were tape recorded and later played back to the panelists in order to stimulate recall of their performances and provide a setting in which immediate reinforcement could be simulated. During the playback the instructor stopped the recordings at points when a panelist was performing well and briefly commented on the merit of the panelist's contribution.

Results of an analysis of covariance support the hypothesis that group discussion competencies of students can be considerably improved by simulating "immediate" reward conditions through the use of stimulated recall techniques. Practice without "immediate" reward conditions had no significant influence on group discussion abilities.

## REFERENCES

Bloom, B. S. Thought processes in lectures and discussion. *J. gen. Educ.*, 1953, **7**, 160–169.

Bloom, B. S., & Broder, Lois J. *Problem-solving processes of college students.* Chicago: Univer. Chicago Press, 1950.

Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

Fouriezos, N. T., Hutt, M. L., & Guetzkow, H. Measurement of self-oriented needs in discussion groups. *J. abnorm. soc. Psychol.,* 1950, **45**, 682–690.

Krasner, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.,* 1958, **55**, 148–170.

Loree, M. R. (Ed.) *Educational psychology.* New York: Ronald, 1959.

Salzinger, K. Experimental manipulation of verbal behavior: A review. *J. gen. Psychol.,* 1959, **61**, 65–94.

Skinner, B. F. *Science and human behavior.* New York: Macmillan, 1953.

Skinner, B. F. Teaching machines. *Science,* 1958, **128**, 969–977.

Wolfe, D. Training. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: John Wiley, 1951. Pp. 1267–1286.

# Part IV

## *Learning: Cognitive Abilities and Processes*

*INQUIRY INTO the nature of learning is a major concern of educational psychology. The next three parts in the present volume will be devoted to this inquiry. Part IV will present seven selections concerning cognitive abilities and processes as these relate to learning. The selections cover the discovery method in learning, concept attainment, problem solving, critical thinking, and creative thinking.*

*Jerome Bruner discusses in the first article the cognitive processes involved in the act of discovery. A professor associated with the Laboratory of Social Relations at Harvard University, Bruner has stirred considerable interest among educators with his recent book,* The Process of Education. *In "The Act of Discovery" he makes a strong case in support of learning by discoveries one makes for oneself.*

*The next article is an example of recent investigations into concept attainment. The research reported here was conducted by Lee Yudin and Solis Kates, both of the University of Massachusetts. In method, the experiment approximates the procedures employed by Bruner et al., as reported in* A Study of Thinking; *in theory, it is based on the work of Inhelder and Piaget.*

*Howard Kendler and Tracy Kendler, the first of New York University and the second of Barnard College, attempt to demonstrate in reading number 21 the possibility of investigating problem solving in a systematic fashion. They report a series of experiments supporting mediational S-R theory as one possible systematic approach.*

*An investigation of observed qualitative differences among students while engaged in solving arithmetic problems is reported in the next selection. The investigators, Herbert J. Klausmeier and Leo J. Loughlin, of the University of Wisconsin and Northern Illinois University, respectively, attempt to clarify factors contributing to the wide range of problem-solving abilities among school children.*

*Reading number 23, "A Concept of Critical Thinking," is a modified version of an article that originally appeared in the Winter, 1962, issue of the* Harvard Educational Review. *The author, Robert H. Ennis, professor of education at Cornell University, defines critical thinking as "the correct assessing of statements." The concept of critical thinking presented in this paper has generated a large-scale research project at Cornell's School of Education.*

*In reading number 24, Donald MacKinnon discusses the nature and nurture of creative talent. MacKinnon is director of the Institute of Personality Assessment and Research at the University of California. The present paper was* *originally given as the Walter Van Dyke Bingham Lecture, for 1962, at Yale University. Findings from research carried on over a period of years at the Institute serve as a basis for MacKinnon's remarks.*

*Part IV concludes with a correlational study of the relationship between intelligence and creative-thinking abilities by this author and Frank B. May, of the University of Wisconsin. May was at Washington State University when the study was published. In it the well-known effect of range of measures on the magnitude of the correlation coefficient is used to emphasize the need for caution in stating relationships between IQ and creative-thinking performance.*

# 19

# The Act of Discovery — *Jerome S. Bruner*

*The importance of active participation by the pupil in the learning process is emphasized in this article. The author suggests that when one discovers knowledge—in contrast to being given knowledge by the teacher—a unique relation is established between that knowledge and its possessor. Discovery involves the manipulation of facts and ideas, and the movement beyond them "to additional new insights." In discussing cognitive factors in the process of discovery, the author hypothesizes effects on the way in which the learner is able to acquire, retrieve, and use knowledge.*

MAIMONIDES, in his *Guide for the Perplexed*[1], speaks of four forms of perfection that men might seek. The first and lowest form is perfection in the acquisition of worldly goods. The great philosopher dismisses such perfection on the ground that the possessions one acquires bear no meaningful relation to the possessor: "A great king may one morning find that there is no difference between him and the lowest person." A second perfection is of the body, its conformation and skills. Its failing is that it does not reflect on what is uniquely human about man: "he could [in any case] not be as strong as a mule." Moral perfection is the third, "the highest degree of excellency in man's character." Of this perfection Maimonides says: "Imagine a person being alone, and having no connection whatever with any other person; all his good moral principles are at rest, they are not required and give man no perfection whatever. These principles are only necessary and

---

Reprinted from *Harvard Educational Review*, 1961, 31, 21–32, by permission of the publisher.

[1] Maimonides, *Guide for the perplexed,* New York: Dover Publications, 1956.

useful when man comes in contact with others." "The fourth kind of perfection is the true perfection of man; the possession of the highest intellectual faculties. . . ." In justification of his assertion, this extraordinary Spanish-Judaic philosopher urges: "Examine the first three kinds of perfection; you will find that if you possess them, they are not your property, but the property of others. . . . But the last kind of perfection is exclusively yours; no one else owns any part of it."

It is a conjecture much like that of Maimonides that leads me to examine the act of discovery in man's intellectual life. For if man's intellectual excellence is the most his own among his perfections, it is also the case that the most uniquely personal of all that he knows is that which he has discovered for himself. What difference does it make, then, that we encourage discovery in the learning of the young? Does it, as Maimonides would say, create a special and unique relation between knowledge possessed and the possessor? And what may such a unique relation do for a man—or for a child, if you will, for our concern is with the education of the young?

The immediate occasion for my concern with discovery—and I do not restrict discovery to the act of finding out something that before was unknown to mankind, but rather include all forms of obtaining knowledge for oneself by the use of one's own mind—the immediate occasion is the work of the various new curriculum projects that have grown up in America during the last six or seven years. For whether one speaks to mathematicians or physicists or historians, one encounters repeatedly an expression of faith in the powerful effects that come from permitting the student to put things together for himself, to be his own discoverer.

First, let it be clear what the act of discovery entails. It is rarely, on the frontier of knowledge or elsewhere, that new facts are "discovered" in the sense of being encountered as Newton suggested in the form of islands of truth in an uncharted sea of ignorance. Or if they appear to be discovered in this way, it is almost always thanks to some happy hypotheses about where to navigate. Discovery, like surprise, favors the well prepared mind. In playing bridge, one is surprised by a hand with no honors in it at all and also by hands that are all in one suit. Yet all hands in bridge are equiprobable: one must know to be surprised. So too in discovery. The history of science is studded with examples of men "finding out" something and not knowing it. I shall operate on the assumption that discovery, whether by a schoolboy going it on his own or by a scientist cultivating the growing edge of his field, is in its essence a matter of rearranging or transforming evidence in such a way that one is enabled to go beyond the evidence so reassembled to additional new insights. It may well be that an additional fact or shred of evidence makes this larger transformation of evidence possible. But it is often not even dependent on new information.

It goes without saying that, left to himself, the child will go about discovering things for himself within limits. It also goes without saying that there are certain forms of child rearing, certain home atmospheres that lead some children to be their own discoverers more than other children. These are both topics of great interest, but I shall not be discussing them. Rather, I should like to confine myself to the consideration of discovery and "finding-out-for-oneself" within an educational setting—specifically the school. Our aim as teachers is to give our student as firm a grasp of a subject as we can, and to make him as autonomous and self-propelled a thinker as we can—one who will go along on his own after formal schooling has ended. I shall return in the end to the question of the kind of classroom and the style of teaching that encourages an attitude of wanting to discover. For purposes of orienting the discussion, however, I would like to make an overly simplified distinction between teaching that takes place in the *expository mode* and teaching that utilizes the *hypothetical mode*. In the former, the decisions concerning the mode and pace and style of exposition are principally determined by the teacher as expositor; the student is the listener. If I can put the matter in terms of structural linguistics, the speaker has a quite different set of decisions to make than the listener: the former has a wide choice of alternatives for structuring, he is anticipating paragraph content while the listener is still intent on the words, he is manipulating the content of the material by various transformations, while the listener is quite unaware of these internal manipulations. In the hypothetical mode, the teacher and the student are in a more cooperative position with respect to what in linguistics would be called "speaker's decisions." The student is not a bench-bound listener, but is taking a part in the formulation and at times may play the principal role in it. He will be aware of alternatives and may even have an "as if" attitude toward these and, as he receives information he may evaluate it as it comes. One cannot describe the process in either mode with great precision as to detail, but I think the foregoing may serve to illustrate what is meant.

Consider now what benefit might be derived from the experience of learning through discoveries that one makes for oneself. I should like to discuss these under four headings: (1) The increase in intellectual potency, (2) the shift from extrinsic to intrinsic rewards, (3) learning the heuristics of discovering, and (4) the aid to memory processing.

## Intellectual Potency

If you will permit me, I would like to consider the difference between subjects in a highly constrained psychological experiment involving a two-choice apparatus. In order to win chips, they must depress a key either on the right or the left side of the machine. A pattern of payoff is designed such that, say, they will be paid off on the right side 70 per cent of the time, on the left 30 per cent, although this detail is not important. What is important is that the payoff sequence is arranged at random, and there is no pattern. I should like to contrast the behavior of subjects who think that there *is* some pattern to be found in the sequence—who think that regularities are discoverable—in contrast to subjects who think that things are happening quite by *chance*. The fomer group adopts what is called an "event-matching" strategy in which the number of responses given to each side is roughly equal to the proportion of times it pays off: in the present case R70 : L30. The group that believes there is no pattern very soon reverts to a much more primitive strategy wherein *all* responses are allocated to the side that has the greater payoff. A little arithmetic will show you that the lazy all-and-none strategy pays off more if indeed the environment is random: namely, they win seventy per cent of the time. The event-matching subjects win about 70% on the 70% payoff side (or 49% of the time there) and 30% of the time on the side that pays off 30% of the time (another 9% for a total take-home wage of 58% in return for their labors of decision). But the world is not always or not even frequently random, and if one analyzes carefully what the event-matchers are doing, it turns out that they are trying out hypotheses one after the other, all of them containing a term such that they distribute bets on the two sides with a frequency to match the actual occurrence of events. If it should turn out that there is a pattern to be discovered, their payoff would become 100%. The other group would go on at the middling rate of 70%.

What has this to do with the subject at hand? For the person to search out and find regularities and relationships in his environment, he must be armed with an expectancy that there will be something to find and, once aroused by expectancy, he must devise ways of searching and finding. One of the chief enemies of such expectancy is the assumption that there is nothing one can find in the environment by way of regularity or relationship. In the experiment just cited, subjects often fall into a habitual attitude that there is either nothing to be found or that they can find a pattern by looking. There is an important sequel in behavior to the two attitudes, and to this I should like to turn now.

We have been conducting a series of experimental studies on a group of some seventy school children over the last four years. The studies have led us to distinguish an interesting dimension of cognitive activity that can be described as ranging from *episodic empiricism* at one end to *cumulative constructionism* at the other. The two attitudes in the choice experiments just cited are illustrative of the extremes of the dimension. I might mention some other illustrations. One of the experiments employs the game of Twenty Questions. A child—in this case he is between 10 and 12—is told that a car has gone off the road and hit a tree. He is to ask questions that can be answered by "yes" or "no" to discover the cause of the accident. After completing the problem, the same task is given him again, though he is told that the accident had a different cause this time. In all, the procedure is repeated four times. Children enjoy playing the game. They also differ quite markedly in the approach or strategy they bring to the task. There are various elements in the strategies employed. In the first place, one may distinguish clearly between two types of questions asked: the one is designed for locating constraints in the problem, constraints that will eventually give shape to an hypothesis; the other is the hypothesis as question. It is the difference between, "Was there anything wrong with the driver?" and "Was the driver rushing to the doctor's office for an appointment and the car got out of control?" There are children who precede hypotheses with efforts to locate constraint and there are those who, to use our local slang, are "pot-shotters," who string out hypotheses non-cumulatively one after the other. A second element of strategy is its connectivity of information gathering: the extent to which questions asked utilize or ignore or violate information previously obtained. The questions asked by children tend to be organized in cycles, each cycle of questions usually being given over to the pursuit of some particular notion. Both within cycles and between cycles one can discern a marked difference on the connectivity of the child's performance. Needless to say, children who employ constraint location as a technique preliminary to the formulation of hypotheses tend to be far more connected in their harvesting of information. Persistence is another feature of strategy, a characteristic compounded of what appear to be two components: a sheer doggedness component, and a persistence that stems from the sequential organization that a child brings to the task. Doggedness is probably just animal spirits or the need for achievement—what has come to be called *n-ach*. Organized persistence is a maneuver for protecting our fragile cognitive apparatus from overload. The child who has flooded himself with disorganized information from unconnected hypotheses will become discouraged and confused sooner than the child who has shown a certain cunning in his strategy of getting information—a cunning whose principal component is the recognition that the value of information is not simply in getting it but in being able to carry it. The persistence of the organized child stems from his knowledge of how to organize questions in cycles, how to summarize things to himself, and the like.

Episodic empiricism is illustrated by information gathering that is unbound by prior constraints, that lacks connectivity, and that is deficient in organizational persistence. The opposite extreme is illustrated by an approach that is characterized by constraint sensitivity, by connective maneuvers, and by organized persistence. Brute persistence seems to be one of those gifts from the gods that make people more exaggeratedly what they are.[2]

Before returning to the issue of discovery and its role in the development of thinking, let me say a word more about the ways in which information may get transformed when the problem solver has actively processed it. There is first of all a pragmatic question: what does it take to get information processed into a form best designed to fit some future use? Take an experiment by Zajonc[3] as a case in point. He gives groups of subjects information of a controlled kind, some groups being told that their task is to transmit the information to others, others that it is merely to be kept in mind. In general, he finds more differentiation and organization of the information received with the intention of being transmitted than there is for information received passively. An active set leads to a transformation related to a task to be performed. The risk, to be sure, is in possible overspecialization of information processing that may lead to such a high degree of specific organization that information is lost for general use.

I would urge now in the spirit of an hypothesis that emphasis upon discovery in learning has precisely the effect upon the learner of leading him to be a constructionist, to organize what he is encountering in a manner not only designed to discover regularity and relatedness, but also to avoid the kind of information drift that fails to keep account of the uses to which information might have to be put. It is, if you will, a necessary condition for learning the variety of techniques of problem solving, of transforming information for better use, indeed for learning how to go about the very task of learning. Practice in discovering for oneself teaches one to acquire information in a way that makes that information more readily viable in problem solving. So goes the hypothesis. It is still in need of testing. But it is an hypothesis of such important human implications that we cannot afford not to test it—and testing will have to be in the schools.

## Intrinsic and Extrinsic Motives

Much of the problem in leading a child to effective cognitive activity is to free him from the immediate control of environmental rewards and punishments. That is to say, learning that starts in response to the rewards of parental or teacher approval or the avoidance of failure can too readily develop a pattern in which the child is seeking cues as to how to conform to what is expected of him. We know from studies of children who tend to be early over-achievers in school that they are likely to be seekers after the "right way to do it" and that their capacity for transforming their learning into viable thought structures tends to be lower than children merely achieving at levels predicted by intelligence tests. Our tests on such children show them to be lower in analytic ability than those who are not conspicuous in overachievement.[4] As we shall see later, they develop rote abilities and depend upon being able to "give back" what is expected rather than to make it into something that relates to the rest of their cognitive life. As Maimonides would say, their learning is not their own.

---

[2] I should also remark in passing that the two extremes also characterize concept attainment strategies as reported in *A study of thinking* by J. S. Bruner et al., New York: John Wiley, 1956. Successive scanning illustrates well what is meant here by episodic empiricism; conservative focussing is an example of cumulative constructionism.

[3] R. B. Zajonc, personal communication, 1957.

The hypothesis that I would propose here is that to the degree that one is able to approach learning as a task of discovering something rather than "learning about" it, to that degree will there be a tendency for the child to carry out his learning activities with the autonomy of self-reward or, more properly, by reward that is discovery itself.

To those of you familiar with the battles of the last half-century in the field of motivation, the above hypothesis will be recognized as controversial. For the classic view of motivation in learning has been, until very recently, couched in terms of a theory of drives and reinforcement: that learning occurred by virtue of the fact that a response produced by a stimulus was followed by the reduction in a primary drive state. The doctrine is greatly extended by the idea of secondary reinforcement: any state associated even remotely with the reduction of a primary drive could also have the effect of producing learning. There has recently appeared a most searching and important criticism of this position, written by Professor Robert White,[5] reviewing the evidence of recently published animal studies, of work in the field of psychoanalysis, and of research on the development of cognitive processes in children. Professor White comes to the conclusion, quite rightly I think, that the drive-reduction model of learning runs counter to too many important phenomena of learning and development to be either regarded as general in its applicability or even correct in its general approach. Let me summarize some of his principal conclusions and explore their applicability to the hypothesis stated above.

> I now propose that we gather the various kinds of behavior just mentioned, all of which have to do with effective interaction with the environment, under the general heading of competence. According to Webster, competence means fitness or ability, and the suggested synonyms include capability, capacity, efficiency, proficiency, and skill. It is therefore a suitable word to describe such things as grasp-

---

[4] J. S. Bruner and A. J. Caron, Cognition, anxiety, and achievement in the preadolescent, *J. educ. Psychol.* (in press).

[5] R. W. White, Motivation reconsidered: The concept of competence, *Psychol. Rev.*, **66**, 1959, 297–333.

> ing and exploring, crawling and walking, attention and perception, language and thinking, manipulating and changing the surroundings, all of which promote an effective—a competent—interaction with the environment. It is true of course, that maturation plays a part in all these developments, but this part is heavily overshadowed by learning in all the more complex accomplishments like speech or skilled manipulation. I shall argue that it is necessary to make competence a motivational concept; there is *competence motivation* as well as competence in its more familiar sense of achieved capacity. The behavior that leads to the building up of effective grasping, handling, and letting go of objects, to take one example, is not random behavior that is produced by an overflow of energy. It is directed, selective, and persistent, and it continues not because it serves primary drives, which indeed it cannot serve until it is almost perfected, but because it satisfies an intrinsic need to deal with the environment.[6]

I am suggesting that there are forms of activity that serve to enlist and develop the competence motive, that serve to make it the driving force behind behavior. I should like to add to White's general premise that the *exercise* of competence motives has the effect of strengthening the degree to which they gain control over behavior and thereby reduce the effects of extrinsic rewards or drive gratification.

The brilliant Russian psychologist Vigotsky[7] characterizes the growth of thought processes as starting with a dialogue of speech and gesture between child and parent; autonomous thinking begins at the stage when the child is first able to internalize these conversations and "run them off" himself. This is a typical sequence in the development of competence. So too in instruction. The narrative of teaching is of the order of the conversation. The next move in the development of competence is the internalization of the narrative and its "rules of generation" so that the child is now capable of running off the narrative on his own. The hypothetical mode in teaching by encouraging the child to participate in "speaker's decisions" speeds this process along. Once internalization has occurred, the child is in a vastly improved position from several obvious points of view—notably that he is able to go beyond the information he has been given to generate additional ideas that can either be checked immediately from experience or can, at least, be used as a basis for formulating reasonable hypotheses. But over and beyond that, the child is now in a position to experience success and failure not as reward and punishment, but as information. For when the task is his own rather than a matter of matching environmental demands, he becomes his own paymaster in a certain measure. Seeking to gain control over his environment, he can now treat success as indicating that he is on the right track, failure as indicating he is on the wrong one.

[6] *Ibid.*, pp. 317–18.
[7] L. S. Vigotsky, *Thinking and speech*, Moscow, 1934.

In the end, this development has the effect of freeing learning from immediate stimulus control. When learning in the short run leads only to pellets of this or that rather than to mastery in the long run, then behavior can be readily "shaped" by extrinsic rewards. When behavior becomes more long-range and competence-oriented, it comes under the control of more complex cognitive structures, plans, and the like, and operates more from the inside out. It is interesting that even Pavlov, whose early account of the learning process was based entirely on a notion of stimulus control of behavior through the conditioning mechanism in which, through contiguity a new conditioned stimulus was substituted for an old unconditioned stimulus by the mechanism of stimulus substitution, that even Pavlov recognized his account as insufficient to deal with higher forms of learning. To supplement the account, he introduced the idea of the "second signalling system," with central importance placed on symbolic systems such as language in mediating and giving shape to mental life. Or as Luria[8] has put it, "the first signal system [is] concerned with directly perceived stimuli, the second with systems of verbal elaboration." Luria, commenting on the importance of the transition from first to second signal system, says: "It would be mistaken to suppose that verbal intercourse with adults merely changes the contents of the child's conscious activity without changing its form. . . . The word has a basic function not only because it indicates a corresponding object in the external world, but also because it abstracts, isolates the necessary signal, generalizes perceived signals and relates them to certain categories; it is this systematization of direct experience that makes the role of the word in the formation of mental processes so exceptionally important."[9]

It is interesting that the final rejection of the universality of the doctrine of reinforcement in direct conditioning came from some of Pavlov's own students. Ivanov-Smolensky[10] and Krasnogorsky[11] published papers showing the manner in which symbolized linguistic messages could take over the place of the unconditioned stimulus and of the unconditioned response (gratification of hunger) in children. In all instances, they speak of these as *replacements* of lower, first-system mental or neural processes by higher order or second-system controls. A strange irony, then, that Russian psychology that gave us the notion of the conditioned response and the assumption that higher order activities are built up out of colligations or structurings of such primitive units, rejected this notion while much of American learning psychology has stayed until quite recently within the early Pavlovian fold (see, for example, a recent article by Spence[12] in the *Harvard Educational Review* or Skinner's treatment of language[13] and the attacks that have been made upon it by linguists such as Chomsky[14] who have become concerned with the relation of language and cognitive activity). What is the more interesting is that Russian pedagogical theory has become deeply influenced by this new trend and is now placing much stress upon the importance of building up a more active symbolical approach to problem solving among children.

---

[8] A. L. Luria, The directive function of speech in development and dissolution, *Word,* **15**, 1959, 341–464.

[9] *Ibid.*, p. 12. For an elaboration of the view expressed by Luria, the reader is referred to the forthcoming translation of L. S. Vigotsky's 1934 book being published by John Wiley and Sons and the Technology Press.

[10] A. G. Ivanov-Smolensky, Concerning the study of the joint activity of the first and second signal systems, *J. of higher nerv. Activ.*, **1**, 1951, 1.

[11] N. D. Krasnogorsky, *Studies of higher nervous activity in animals and in man,* Vol. **I**, Moscow, 1954.

To sum up the matter of the control of learning, then, I am proposing that the degree to which competence or mastery motives come to control behavior, to that degree the role of reinforcement or "extrinsic pleasure" wanes in shaping behavior. The child comes to manipulate his environment more actively and achieves his gratification from coping with problems. Symbolic modes of representing and transforming the environment arise and the importance of stimulus-response-reward sequences declines. To use the metaphor that David Riesman developed in a quite different context, mental life moves from a state of outer-directed in which the fortuity of stimuli and reinforcement are crucial to a state of inner-directedness in which the growth and maintenance of mastery become central and dominant.

## Learning the Heuristics of Discovery

Lincoln Steffens,[15] reflecting in his *Autobiography* on his under graduate education at Berkeley, comments that his schooling was overly specialized on learning about the known and that too little attention was given to the task of finding out about what was not known. But how does one train a student in the techniques of discovery? Again I would like to offer some hypotheses. There are many ways of coming to the arts of inquiry. One of them is by careful study of its formalization in logic, statistics, mathematics, and the like. If a person is going to pursue inquiry as a way of life, particularly in the sciences, certainly such study is essential. Yet, whoever has taught kindergarten and the early primary grades or has had graduate students working with him on their theses—I choose the two extremes for they are both periods of intense inquiry—knows that an understanding of the formal aspect of inquiry is not sufficient. There appear to be, rather, a series of activities and attitudes, some directly related to a particular subject and some of them fairly generalized, that go with inquiry and research. These have to do with the *process* of trying to find out something and while they provide no guarantee that the *product* will be any *great* discovery, their absence is likely to lead to awkwardness or aridity or confusion. How difficult it is to describe these matters—the heuristics of inquiry. There is one set of attitudes or ways of doing that has to do with sensing the relevance of variables—how to avoid getting stuck with edge effects and getting instead to the big sources of variance. Partly this gift comes from intuitive familiarity with a range of phenomena, sheer "knowing the stuff." But it also comes out of a sense of what things among an assemble of things "smell right" in the sense of being of the right order of magnitude or scope or severity.

---

[12] K. W. Spence, The relation of learning theory to the technique of education, *Harvard educ. rev.*, 29, 1959, 84–95.

[13] B. F. Skinner, *Verbal behavior*, New York: Appleton-Century-Crofts, 1957.

[14] N. Chomsky, *Syntactic structure*, The Hague, The Netherlands: Mouton & Co., 1957.

[15] L. Steffens. *Autobiography of Lincoln Steffens*, New York: Harcourt, Brace, 1931.

The English philosopher Weldon describes problem solving in an interesting and picturesque way. He distinguishes between difficulties, puzzles, and problems. We solve a problem or make a discovery when we impose a puzzle form onto a difficulty that converts it into a problem that can be solved in such a way that it gets us where we want to be. That is to say, we recast the difficulty into a form that we know how to work with, then work it. Much of what we speak of as discovery consists of knowing how to impose what kind of form on various kinds of difficulties. A small part but a crucial part of discovery of the highest order is to invent and develop models or "puzzle forms" that can be imposed on difficulties with good effect. It is in this area that the truly powerful mind shines. But it is interesting to what degree perfectly ordinary people can, given the benefit of instruction, construct quite interesting and what, a century ago, would have been considered greatly original models.

Now to the hypothesis. It is my hunch that it is only through the exercise of problem solving and the effort of discovery that one learns the working heuristic of discovery, and the more one has practice, the more likely is one to generalize what one has learned into a style of problem solving or inquiry that serves for any kind of task one may encounter—or almost any kind of task. I think the matter is self-evident, but what is unclear is what kinds of training and teaching produce the best effects. How do we teach a child to, say, cut his losses but at the same time be persistent in trying out an idea; to risk forming an early hunch without at the same time formulating one *so* early and with so little evidence as to be stuck with it waiting for appropriate evidence to materialize; to pose good testable guesses that are neither too brittle nor to sinuously incorrigible; etc., etc. Practice in inquiry, in trying to figure out things for oneself is indeed what is needed, but in what form? Of only one thing I am convinced. I have never seen anybody improve in the art and technique of inquiry by any means other than engaging in inquiry.

## Conservation of Memory

I should like to take what some psychologists might consider a rather drastic view of the memory process. It is a view that in large measure derives from the work of my colleague, Professor George Miller.[16] Its first premise is that the principal problem of human memory is not storage, but retrieval. In spite of the biological unlikeliness of it, we seem to be able to store a huge quantity of information—perhaps not a full tape recording, though at times it seems we even do that, but a great sufficiency of impressions. We may infer this from the fact that recognition (i.e., recall with the aid of maximum prompts) is so extraordinarily good in human beings—particularly in comparison with spontaneous recall where, so to speak, we must get out stored information without external aids or prompts. The key to retrieval is organization or, in even simpler terms, knowing where to find information and how to get there.

Let me illustrate the point with a simple experiment. We present pairs of words to twelve-year-old children. One group is simply told to remember the pairs, that they will be asked to repeat them later. Another is told to remember them by producing a word or idea that will tie the pair together in a way that will make sense to them. A third group is given the mediators used by the second group when presented with the pairs to aid them in tying the pairs into working units. The word pairs include such juxtapositions as "chair-forest," "sidewalk-square," and the like. One can distinguish three styles of mediators and children can be scaled in terms of their relative preference for each: *generic mediation* in which a pair is tied together by a superordinate idea: "chair and forest are both made of wood"; *thematic mediation* in which the two terms are imbedded in a theme or little story: "the lost child sat on a chair in the middle of the forest"; and *part-whole mediation* where "chairs are made from trees in the forest" is typical. Now, the chief result, as you would all predict, is that children who provide their own mediators do best—indeed, one time through a set of thirty pairs, they recover up to 95% of the second words when presented with the first ones of the pairs, whereas the uninstructed children reach a maximum of less than 50% recovered. Interestingly enough, children do best in recovering materials tied together by the form of mediator they most often use.

One can cite a myriad of findings to indicate that any organization of information that reduces the aggregate complexity of material by imbedding it into a cognitive structure a person has constructed will make that material more accessible for retrieval. In short, we may say that the process of memory, looked at from the retrieval side, is also a process of problem solving: how can material be "placed" in memory so that it can be got on demand?

---

[16] G. A. Miller, The magical number seven, plus or minus two, *Psychol. Rev.*, 63, 1956, 81–97.

We can take as a point of departure the example of the children who developed their own technique for relating the members of each word pair. You will recall that they did better than the children who were given by exposition the mediators they had developed. Let me suggest that in general, material that is organized in terms of a person's own interests and cognitive structures is material that has the best chance of being accessible in memory. That is to say, it is more likely to be placed along routes that are connected to one's own ways of intellectual travel.

In sum, the very attitudes and activities that characterize "figuring out" or "discovering" things for oneself also seem to have the effect of making material more readily accessible in memory.

# 20

# Concept Attainment and Adolescent Development — *Lee Yudin and Solis L. Kates*

*Recently considerable effort has been devoted to the investigation of efficiency and strategy in concept attainment. The investigation reported here exemplifies the type of research being carried out along these lines. The experimenters test two hypotheses derived from the developmental theory of Inhelder and Piaget. Using groups of 12-year-old, 14-year-old, and 16-year-old adolescents as subjects, efficiency and strategy in concept attainment are related to age and experimental method of presenting instances of the concept. Results confirm the description of adolescent development advanced by Inhelder and Piaget. Method of presentation (i.e., presentation of positive or negative instances of a concept, both giving equal information) did not produce a significant differential effect.*

THE PURPOSE of this study was to determine if efficiency and strategy in concept attainment are related to age and to experimental method of presenting instances of the concept. The developmental theory of Inhelder and Piaget (1958) indicated that, as adolescents grow older, they develop greater facility in generalizing from the data at hand, in combining operations such as classification and seriation, in formulating and manipulating propositions, and in reasoning in a hypothetico-deductive fashion. Further experimental evidence is necessary to demonstrate that more efficient

Reprinted from *Journal of Educational Psychology*, 1963, 54, 177–182 by permission of the authors and the American Psychological Association.

intellectual operations are used by 14-year-old adolescents as compared with 12-year-old adolescents, as suggested by these investigators.

Hovland and Weiss (1953) have explored whether negative instances or positive instances that transmit equal information are equally effective in concept attainment. Concept attainment, they found, was more efficient with positive instances than with negative instances. A pertinent question to be investigated is whether this finding applies to concept attainment by adolescents under optimal conditions of presenting information. Cahill and Hovland (1960) have demonstrated that a simultaneous successive method for concept attainment was superior to a successive method. One of their measures, the perceptual error—a guess about the concept that is incompatible with the information in the current instance—was found to be highly correlated with concept attainment efficiency.

In their concept attainment studies, Bruner, Goodnow, and Austin (1956) controlled the presentation of the instances and identified the instances as exemplars or nonexemplars of the concept. Their subjects guessed the concept after each instance and tended to follow ideal rules for maintaining or changing their guesses. Upon the basis of the Bruner et al. research, we have developed two methods of examining the sequence of guesses, the consistent and the ideal strategies. The consistent strategy indicates that the guess is maintained or changed when the current instance is confirming or infirming. The ideal strategy is one where the guess is maintained or changed in accordance with the information presented by the current and all previous instances.

Perceptual error, consistent strategy, and ideal strategy, in that order, appear to be measures representing increasing maturity of cognitive operations. Perceptual error suggests an inability to deal with information given directly and perceptually. Consistent strategy shows an ability to deal with information by maintaining or changing guesses but without the systematic integration of previous information. Ideal strategy includes the ability defined by the consistent strategy and the additional ability to integrate systematically previous information.

The hypotheses are as follows:

1. Twelve-year-old subjects will require significantly more instances for concept attainment, will have significantly fewer consistent and ideal strategies, and will have significantly more perceptual errors than 14- and 16-year-old subjects. The 14- and the 16-year-old subjects will not differ significantly on these variables.

2. Concept attainment on positive instances will require significantly fewer instances, will have significantly more consistent and ideal strategies, and will have significantly fewer perceptual errors than concept attainment on negative instances presenting equal information.

## Method

### Subjects

There were 40 boys at the 12-year level, 40 boys at the 14-year level, and 40 boys at the 16-year level, making a total of 120 subjects. These boys were within 6 months of their birthdays. Subjects were from different public schools in western Massachusetts but were equally represented in the experimental groups to be described below.

### Test Materials

All the subjects were administered two practice problems and six experimental problems. Each problem consisted of 13 2 × 2 inch slides. After the first slide, each successive cycle of 3 slides provided sufficient information to solve the problem. The problem was solved when the correct answer was offered twice consecutively without any subsequent incorrect responses.

The first slide in each problem presented one positive instance of the concept, the focus instance. The remaining 12 slides for any problem included, on each slide, the focus instance and one other instance. For example, on the slides of the positive series, the focus instance and one other positive instance of the concept appeared. On the slides of the negative series, the focus instance and a negative instance were shown.

Each instance showed a combination of four attributes: (*a*) one of three forms (circles, crosses, or squares), (*b*) one of three colors for the forms (red, black, or green), (*c*) one of three numbers of forms (one, two, or three), and (*d*) one of three borders (one, two, or three lines surrounding the forms). Several examples of instances are: (*a*) one red circle, two borders; (*b*) two black crosses, three borders; (*c*) three green squares, one border; (*d*) two black circles, one border.

The concepts to be solved were the same for the corresponding problems in the positive and negative series. For example, in both positive and negative series, the concept to be attained in the first problem was the color black; the focus instance in the first problem of the two series was three black squares and two borders. The second problem for both positive and negative series required all the subjects to discover that the concept was one form given the focus instance of one red cross and two borders. The same conditions obtained for the other four experimental problems (Yudin, 1962).

The instructions indicated that the correct concept in each problem was a single attribute. As mentioned above, the single attribute black was the concept for the first problem and the single attribute of one form was the concept for the second problem. Black crosses or one square could not be correct because these responses contained two attributes.

Equal information was presented by the corresponding slides of the corresponding problems in the positive and negative series. The second slide of the particular problem, whether in the positive or negative series, eliminated the same attribute. The third slide in this problem eliminated an additional but same attribute. For example, the second slide of the first problem, whether in the positive or negative series, eliminated number; the third slide of this problem eliminated form. The experimental problems, the instances contained on the 13 slides for each problem, and the attributes eliminated by each slide, are available in Yudin (1962).

**Procedure**

The 120 subjects were divided into four experimental groups of 30 subjects each. The 30 subjects in any one experimental group included 10 12-year-olds, 10 14-year-olds, and 10 16-year-olds. The 12-year-old subjects in all the four experimental groups were matched on intelligence and educational achievement; similarly, the 14- and 16-year-old subjects in all four groups were matched on intelligence and educational achievement.

The first experimental group was shown positive instances in their six experimental problems. The second group was shown positive instances in the first three problems and negative instances in the second three problems. The third group solved the first three problems given negative instances and the second three problems given positive instances. The six problems solved by the fourth group contained negative instances.

The problems were presented to groups of five subjects. The attributes of the instances were pointed out and examples of instances were demonstrated. The full instructions are given in Yudin (1962). The experimental problems were administered when the subjects in each group had successfully solved the practice problems. All subjects solved the same six experimental problems.

We emphasized that any one attribute could be the correct concept for any problem. By this instruction, we tried to break any possible set that may have been established by the practice problems or by the preceding experimental problems. After each slide was exposed for 10 seconds, the subjects wrote their guesses about the concept for that particular problem. Each slide was exposed only once. The subjects covered their guesses with a sheet of paper provided by the experimenter.

For the number of instances, we noted the slide at which the correct concept was offered in each problem. The consistent strategies were determined by counting the number of times a previous guess was maintained when the new instance confirmed it or was changed to accord with the new instance when it was infirmed. For the number of ideal strategies, we counted the number of guesses that was maintained or changed when indicated by the new instance and by all previous instances. A perceptual error was counted when the guess was not in accord with the information given by the immediately preceding slide.

## Results

The three age groups did not differ significantly on intelligence, all groups having an average IQ of 101 on the Otis Beta test. The four experimental groups did not differ significantly on age, intelligence, and school achievement.

### Age and Concept Attainment (First Hypothesis)

There was a significant difference in the number of instances required for concept attainment when the three age groups were compared (see Table 1). Further analysis by the Duncan range test (Edwards, 1960) showed that the 12-year-old subjects required significantly more instances than either the 14- or 16-year-old subjects. The differences between the mean number of instances for the 12-year-old group (54.8) and the 14-year- (44.1) and the 16-year- (42.9) old groups were 10.7 and 11.9, respectively (see Table 2).

TABLE 1. Analysis of Variance Table for Dependent Variables

| | | Number of Instances | | Consistent Strategies | | Ideal Strategies | | Perceptual Errors | |
|---|---|---|---|---|---|---|---|---|---|
| Source | *df* | SS | *F* | SS | *F* | SS | *F* | SS | *F* |
| Age (A) | 2 | 3425 | 5.9°° | 1099 | 6.2°° | 4838 | 8.5°° | 500 | 4.2° |
| Presentation (B) | 3 | 527 | .6 | 189 | .7 | 1067 | 1.2 | 701 | 3.9° |
| A × B | 6 | 1763 | 1.0 | 327 | .6 | 2089 | 1.2 | 730 | 2.0 |
| Subjects /A×B | 108 | 31499 | | 9521 | | 30601 | | 6465 | |
| Total | 119 | 37215 | | 11136 | | 38596 | | 8396 | |

° $p \leq .05$.
°° $p \leq .01$.

TABLE 2. Means and Standard Deviations for Age Groups on Dependent Variables

| | Number of Instances | | Consistent Strategies | | Ideal Strategies | | Perceptual Errors | |
|---|---|---|---|---|---|---|---|---|
| Age Group | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| 12 years | 54.8 | 16.4 | 56.3 | 9.8 | 31.0 | 14.5 | 11.5 | 7.3 |
| 14 years | 44.1 | 17.0 | 61.7 | 9.3 | 42.7 | 17.7 | 7.8 | 7.5 |
| 16 years | 42.9 | 15.2 | 63.4 | 7.3 | 45.7 | 13.8 | 6.8 | 7.3 |

These differences exceeded the values of 7.2 and 7.7 required for significance at the .05 level. The difference (1.2) between the 14- and 16-year-old subjects was not significant for instances.

The results of the analysis of variance indicated that the main effect for the three age groups was significant on consistent strategies (see Table 1). Duncan range comparisons showed that the mean number of consistent strategies for the 12-year-old group (56.3) was significantly less than the mean number for the 14-year-old group (61.7) and the mean number for the 16-year-old group (63.4) where differences greater than 4.2 and 4.4, respectively, are significant at the .05 level (see Table 2). There was no significant difference between 14- and 16-year-old subjects on consistent strategies.

On the variable of ideal strategies, there was a significant finding when the three age groups were compared (see Table 1). Duncan range comparisons revealed that the differences in ideal strategies of 11.7 and 14.7 between 12-year-old (31.0) and 14-year-old (42.7) subjects and 12-year-old and 16-year-old (45.7) subjects, respectively, exceeded the differences necessary for significance at the .05 level (7.5 and 8.0) (see Table 2). No significant difference was found between the 14- and the 16-year-old subjects on ideal strategies.

A significant finding for perceptual errors indicated that the three age groups were different (see Table 1). The respective differences of 3.7 and 4.7 between the means for 12-year-olds (11.5) and 14-year-olds (7.8) and 12-year-olds and 16-year-olds (6.8) were significant where differences of 3.3 and 3.5 are required for significance at the .05 level. There was no significant difference between 14- and 16-year-old subjects on perceptual error.

These results supported all parts of the first hypothesis stating that 12-year-old subjects would be significantly inferior on all measures to 14- and 16-year-old subjects who would not be different.

### Method of Presentation (Second Hypothesis)

There were no significant differences on the number of instances, and on consistent and ideal strategies when performance on the positive series was compared with the negatives series (see Tables 1 and 3). On the variable of perceptual errors, there was a significant finding (see Table 1). However, Duncan range analyses indicated that the number of perceptual errors made on the positive series (5.1) was not significantly different from the number of perceptual errors on the negative series (8.9; see Table 3). A difference of 4.2 was necessary for significance at the .05 level. Parenthetically, the significant finding in the analysis of variance was due to the number of perceptual errors (12.2) made in the negative-positive series. The second hypothesis was not confirmed; there were no differences on instances, consistent and ideal strategies, and on perceptual errors when the positive series was compared with the negative series.

TABLE 3. Means and Standard Deviations for Method of Presentation on Dependent Variables

| Method of Presentation | Number of Instances | | Consistent Strategies | | Ideal Strategies | | Perceptual Errors | |
|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| 6 positive | 43.3 | 17.5 | 61.4 | 10.5 | 41.9 | 17.7 | 5.1 | 5.8 |
| 3 positive plus 3 negative | 48.8 | 17.7 | 61.2 | 9.2 | 37.9 | 17.8 | 8.3 | 9.2 |
| 3 negative plus 3 positive | 49.6 | 13.1 | 58.3 | 8.5 | 39.3 | 11.9 | 12.2 | 6.7 |
| 6 negative | 46.3 | 16.1 | 61.0 | 7.1 | 40.1 | 15.6 | 8.9 | 7.3 |

## Discussion

### Age and Concept Attainment

The first hypothesis was supported by the results; the 14- and the 16-year-old subjects were significantly superior to the 12-year-old subjects in concept attainment efficiency. The strategies employed by the 14- and the 16-year-old subjects were also superior to those of the 12-year-old subjects. The older groups did not differ either in efficiency or in strategies. These results confirm the description of adolescent development advanced by Inhelder and Piaget (1958). They suggest that the stage of formal reasoning operations begins at about 12 years and reaches equilibrium at about 14–16 years of age. At these latter ages, Inhelder and Piaget (1958) indicate that subjects succeed more often than 12-year-old subjects in solving problems demanding formal operations.

In the conceptual problems of this research, two rules for transforming information were used. First, positive instances have the correct attribute and second, negative instances do not have the correct attribute. Subjects had to infer that any attribute that did not appear in the positive instances would not be the correct concept. Some subjects had to make the additional inference that any attribute present in both focus and negative instances would not be the correct concept. In addition to these mental operations, the subjects had to remember the information supplied by the former instances. Our results then demonstrate that the 14- and the 16-year-old subjects are more capable in these mental operations than the 12-year-old subjects.

The results of our strategy variables can help us to determine how the 12-year-old subjects were inferior to the 14- and the 16-year-old subjects. Since the 12-year-old subjects made significantly more perceptual errors, it is evident that they are less efficient than the 14- and the 16-year-old subjects in dealing with information given directly and perceptually. Twelve-year-old subjects were more liable to specify, as the correct concept, an attribute that did not appear in the currently shown positive instance or to specify an attribute that did appear in the currently shown negative instance. The proper application of the two rules for transforming information presented by positive and negative instances was more difficult for the 12-year-olds than for the 14- and the 16-year-olds.

The 14- and the 16-year-old subjects showed significantly more consistent and ideal strategies than the 12-year-old subjects. Since the ideal strategies are counted as part of the consistent strategies and thus partially overlap, we will not separate these strategies in our discussion. The following are the inferences based upon the results with consistent and ideal strategies: (*a*) the older subjects were more capable of remembering their previous guesses to check them with the current instances; (*b*) they more readily maintained their guesses when the current instances were confirming or more readily changed them when the current instances were not confirming; (*c*) in the ideal strategies, they more readily maintained or changed these guesses to accord with all the previous instances.

## Method of Presentation

The results did not support the second hypothesis and demonstrated that positive or negative series, both presenting equal information, made very little difference in the number of instances required or in the strategies pursued for solution. In this experiment, unlike the Hovland and Weiss (1953) study, the negative and the positive instances were always presented with the focus (positive) instance. For the subjects who worked with the negative series, the focus instance was shown with each negative instance on each slide. On each slide of the negative series, the subjects were thus aware of the attributes that could be correct (in the focus instance) and of the one attribute that was not correct (in the negative instance). This presentation of negative instances in conjunction with the same focus instance may have resulted in relatively greater efficiency of concept attainment. Possibly, Smoke's (1932) original suggestion that negative instances might possibly contribute to efficiency of concept attainment holds only when each negative instance is shown together with the same positive instance of the concept. Because there were no redundant instances in each cycle, a more systematic elimination of incorrect attributes was possible without the added strain of dealing with redundant information. Finally, single attributes were the correct concepts, possibly reducing the level of difficulty in assimilating negative information to that of positive information.

## REFERENCES

Bruner, J. S., Goodnow, Jacqueline, J., & Austin, G. A. *A study of thinking.* New York: John Wiley, 1956.

Cahill, H. E., & Hovland, C. I. The role of memory in the acquisition of concepts. *J. exp. Psychol.,* 1960, 59, 137–144.

Edwards, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart & Winston, 1960.

Hovland, C. I., & Weiss, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.,* 1953, 45, 175–182.

Inhelder, Bärbel, & Piaget, J. *The growth of logical thinking.* New York: Basic Books, 1958.

Smoke, K. L. An objective study of concept formation. *Psychol. Monogr.,* 1932, 42, No. 4 (Whole No. 191).

Yudin, L. W. A developmental study of concept attainment. Unpublished master's thesis, Univer. of Massachusetts, 1962.

# 21

# Vertical and Horizontal Processes in Problem Solving — *Howard H. Kendler and Tracy S. Kendler*

*There are several ways of investigating human problem-solving behavior. This paper begins with a discussion of some of the methodological difficulties encountered in problem-solving research. Next, a prethoretical model ("an informal conception that operates as an analogy") of problem solving is presented. This model (a mediational S-R model) is then employed as a directive and stimulus to research on problem solving. A review of the research indicates that, ". . . in their early development, children tend to respond in a manner consistent with a single unit S-R theory. With age, they develop a tendency to respond in a mediational manner." A discussion of possible applications of a mediational S-R model to classroom learning situations should prove interesting.*

THE PRESENT paper is concerned with *an* approach—and not *the* approach—to the universally appealing but nevertheless unpopular research area of problem solving. Problems of problem solving have proved to be particularly refractory to psychologists. More often than not the uncommon researcher with the temerity to attack some aspect of reasoning retreats to more secure and conventional problems when he discovers that his sorties fail to achieve any impressive victory. As a result the literature of problem

Reprinted from *Psychological Review*, 1962, 69, 1–16, by permission of the senior author and the American Psychological Association.

solving is almost chaotic because it is so heavily sprinkled with isolated bits of information (Duncan, 1959).

Perhaps the present stage of development of psychology does not justify the strategy of investigating such a complex phenomenon. Fortunately, or not, science has no built-in traffic lights to inform investigators when to proceed. It may be a risky and potentially unfruitful gambit to investigate problem solving but then again it may not be. In addition to the intrinsic interest of the area it does offer a challenge to those psychologists who are interested in testing the generality of any set of theoretical principles stemming from other areas of behavior (e.g., learning, perception).

This paper initially will make fleeting references to some methodological problems with which a researcher in the field of reasoning must contend. Then a simple pretheoretical model of problem solving will be described, followed by a report of research which the model generated, and which in turn is shaping the model itself.

## Methodological Problems in Problem Solving Research

Anybody who does research is—or should be—aware that every decision he makes cannot be justified by facts or logic. Some decisions must be made on the basis of personal intuition. This is particularly true for the researcher in problem solving who must make three strategic decisions which cannot help but have profound influences on his research and the ideas they generate (Kendler, 1961). These decisions, which are not completely independent, are related to the place of problem solving in psychology, the use of complex or simple experimental tasks, and the selection of a pretheoretical model to guide research. Considering the volitional nature of these problems, as well as the current status of psychological knowledge, it would be both inappropriate and erroneous to consider these methodological problems as offering only one sensible alternative. Adopting this point of view would do much to minimize the needless disputation that seems to perennially surround matters of research strategy.

Accepting the principle that a basic research strategy is not simply an outgrowth of logical and factual considerations does not reduce one to making decisions in either a haphazard or random manner. A given strategy can be adopted on the basis of rational considerations as long as it is realized that other reasonable attitudes might lead to the adoption of different decisions.

The history of problem solving in particular and psychology in general suggests that problem solving can best be conceptualized not as a basic psychological process, but instead as one that reflects the interaction of more fundamental processes (e.g., learning, perception, and motivation).

If problem solving is not viewed as a unitary process, how is an appropriate experimental situation selected to investigate it? One possibility is that a problem can be selected from a "true life" situation such as trouble-shooting electronic equipment. Or problems can be invented (Duncker, 1945; Maier, 1930) that capture the flavor, if only partially, of problems we meet in everyday life.

A more analytical approach can be taken to the selection of an experimental situation to investigate problem solving. If problem solving is compounded of elementary behavioral processes, then it may be more strategic to devise some simple problems in which the relationships of fundamental psychological mechanisms to problem solving are highlighted. That is, tasks should be devised not to duplicate or imitate everyday problems, but instead to isolate and magnify the basic mechanisms that operate in such complex tasks.

This analytical approach which is favored by the authors suffers from one major drawback. How is it possible to know the basic mechanisms of problem solving prior to their discovery? Obviously, excepting divination, there is no method. But this does not prevent the analytical approach from operating. The researcher can prejudge theoretical issues by formulating a model of what he guesses problem solving to be like. The model can guide the investigator in selecting the hypotheses to test, as well as the experimental situations in which to test them.

This brings us to the third and most important decision a problem solving researcher has to make: his choice of a pretheoretical model (Koch, 1959). A pretheoretical model is not equivalent to a theory. The criterion of validity cannot properly be applied to it because essentially a pretheoretical model is an informal conception that operates as an analogy (Lachman, 1960). It is conceivable that different models (e.g., learning, perception, information theory) can all lead to fruitful and valid theories of problem solving.

Psychologists have many possibilities from which to choose their model. These models can be conveniently divided into two main categories: the empirical model that springs primarily from experimental data, and the formal model that is usually generated by mathematical or logical systems. Among the empirical models that have achieved some acceptance are those that are based on introspective findings (e.g., the four successive stage model of "preparation," "incubation," "inspiration," and finally "verification"), the facts of perception, and those of learning. Some formal models used are those dependent upon stochastic models, game theory, and the operation of computers.

The present authors adopted an S-R learning pretheoretical model. The decision no doubt was influenced by professional training and past research efforts. But other considerations entered. For the past 4 decades S-R learning psychologists have probably been the most active experimental and theoretical group in psychology. To some, if not a large, extent this can be attributed to the fruitful and cleansing effect S-R language has upon designing, reporting, and interpreting research. S-R language forces the psychologist to focus his attention on objectively defined environmental and behavior variables and thus encourages the collection of data and the testing of ideas. The efforts of S-R learning psychologists have supplied a host of facts, concepts, and hypotheses that can be exploited in an exploratory excursion into the realm of problem solving.

The facts and theories of learning, however, do not spontaneously coalesce to form a model that can guide research in problem solving. Some selection must be made. S-R learning theory does not represent a single organized formulation. Anyone who is familiar with the systematic orientations of Hull (1952), Guthrie (1952), Spence (1956), and Skinner (1953) is aware of this. Many of these systematic differences, however, become attenuated and some even disappear when viewed from the distance of problem solving behavior. It is possible and perhaps even profitable to develop a learning model for problem solving that ignores many of the points of disagreement among S-R theories.

Much of the objection to S-R language stems from the apparent discrepancy between active, flowing behavior and the inert, static, single S-R association. Using S-R language does not mean that complex behavior *actually* consists of S-R connections. After analyzing the concept of light, Toulmin (1953) concludes: "We do not *find* light atomized into individual rays: we *represent* it as consisting of such rays" (p. 29). Applying the same idea to the concept of the S-R association: "We do not *find* behavior atomized into individual S-R associations: we *represent* it as consisting of such S-R associations." The concept of the S-R association, therefore, must be judged not in terms of its ability to provide a clear image of behavior, but rather in its capacity to represent the facts of behavior.

## Pretheoretical Model of Problem Solving

An S-R model needs to represent two important characteristics of problem solving behavior. These characteristics are behavior is continuous, and at any one time behavior consists of several habits. The terms "horizontal" and "vertical" are used to refer to these processes; horizontal to the continuity of behavior against the dimension of time, and vertical to the assumption that independent levels of behavior (i.e., S-R units) occur simultaneously.

The assumption that S-R associations do not occur in isolation, but instead are linked together to form integrated, continuous behavior goes back many years (e.g., Watson, 1913). Today the process is most commonly referred to as chaining. Skinner (1953) and his associates have developed powerful techniques that shape behavior into long, complicated chains. The mass of data they have collected suggests important principles governing habit chaining. There is little doubt that when their quasitheoretical system is exploited fully with autoinstructional devices that important insights into problem solving behavior will emerge, particularly in relation to how an added bit of knowledge can trigger problem solution. The kind of chaining with which the Skinnerians have dealt (i.e., adding new S-R units to an already functioning chain) does not exhaust all the problems associated with the horizontal processes of problem solving. Of particular importance to problem solving is the *spontaneous* integration of separate habits which occurs when an organism infers the consequences of combining previously independent S-R units. This kind of chaining was investigated in Kohler's (1925) classical studies of insight and in the more controlled reasoning experiments of Maier (1930). More recently the authors (Kendler & Kendler, 1956, 1961; Kendler, Kendler, Pliskoff, & D'Amato, 1958) have tried to identify some of the important variables that enable children to combine separate experiences in order to solve an inference-type problem. Much of the research reported in this paper will be concerned with how mediated stimulus and response events aid in the formation of problem solving chains.

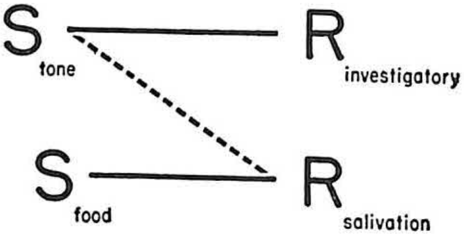


FIG. 1. *An S-R representation of classical conditioning.*

The assumption of vertical processes, i.e., the organism responds several different ways at any one time, is also not a novel one. Every psychologist is aware that organisms make several different responses simultaneously, although typically only one is attended to. Sometimes the different responses are interrelated, as is the case between the heart and respiration rates of a fearful organism. In other cases the different responses are independent, e.g., a person's conversation is uninfluenced by his tugging at his ear lobe. The best laboratory example of vertical processes, and one that has much relevance to problem solving, in shown in Figure 1. Those familiar with introductory psychology textbooks will recognize this diagram as representing classical conditioning. Notice that the two solid lines indicate independent

S-R units which are operating simultaneously. One is the tone that initiates the "investigatory" response, and the other is the food which elicits salivation. Initially these two associations operate in a *parallel* fashion, but as a result of their simultaneous occurrence an *interaction* takes place which is expressed by the broken line representing the acquired conditioned response.

Obviously the brief reference to horizontal and vertical processes in which it is assumed fundamental S-R principles operate (e.g., discrimination, generalization, etc.) presents at best the barest skeleton of a model of problem solving. It needs the flesh and skin of experimental facts to give it solidity and theoretical principles to clothe it in scientific respectability. Let us now review some of the progress that has been made in this direction.

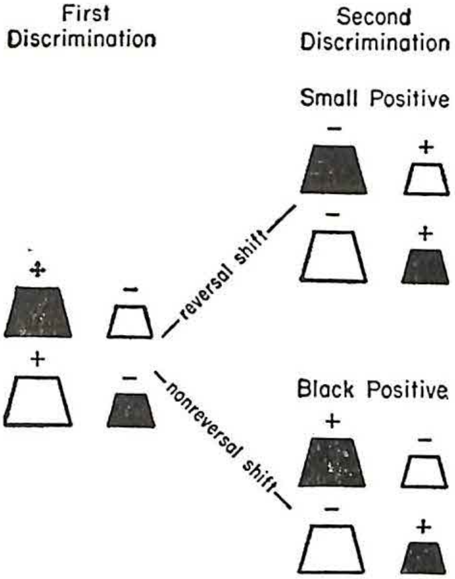


FIG. 2. *Examples of a reversal and a nonreversal shift.*

## Concept Learning and Utilization

Although the primitive model just described fails to generate any research by itself, it does suggest that individual experiments cannot be directed at *problem solving in its entirety.* There are too many aspects to this phenomenon. The researcher, in designing an experiment, must scan the entire problem solving process and then focus upon that segment that promises to yield fruitful results and is also amenable to investigation.

For reasons that will become evident, it was decided to compare reversal and nonreversal shifts in a simple concept learning task. Figure 2 characterizes each kind of shift by showing a *simplified* version of an experimental situation used with children. The stimuli (cups) for their first discrimination differed simultaneously on two dimensions (size and brightness). The subject is rewarded for responses to one dimension (e.g., large cup is positive, small cup is negative). The other dimension is irrelevant. After learning the first discrimination, the subject is forced to shift to another response. In a reversal shift the subject is required to respond to the same dimension on which he was originally trained, but his overt choice has to be reversed, e.g., he has to shift from a *large* cup to a *small* one. For a nonreversal shift the previously irrelevant dimension becomes relevant, e.g., black becomes positive after large had been positive.

Buss (1953) reported that college students executed a reversal shift more rapidly than a nonreversal shift. He attributed this superiority to the intermittent reinforcements that retard the progress of a nonreversal shift. For example, in Figure 2,[1] when a subject is making a nonreversal shift from large positive to black positive, he is reinforced when choosing the large black cup in preference to the small white cup. This fortuitous reinforcement of the choice of the large cup helps maintain the size discrimination and hence retards the learning of the brightness discrimination. The reversal shift group, on the other hand, receives no reinforcement of the previously correct responses, since they are 100% *non*reinforced.

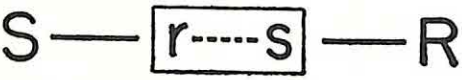


FIG. 3. *A schematic representation of the mediational hypothesis.*

This analysis is at best incomplete. The work of Kendler and Vineberg (1954) suggested that adult human concept learning cannot be represented adequately by a single-unit S-R theory in which the external stimulus is directly connected to the overt response. Instead, a mediational mechanism (see Figure 3) is required which assumes that the external stimulus evokes an implicit response which produces an implicit cue that is connected to the overt response.

It would be useful to digress for a moment to comment about the epistemological status of these inferred stimulus and response events which are enclosed in the rectangle to emphasize their hypothetical character. Although not directly observable, they are "tied to" environmental and behavioral events. The basic assumption of the mediational hypothesis, at least for the time being, is that the implicit stimulus and response events obey the same principles that operate in observable S-R relationships.

---

[1] The purpose of Figure 2 is to clarify the meaning of both a reversal and nonreversal shift. It would be misleading to believe that it represents *exactly* the methodology of "reversal-nonreversal" studies reported in this paper. For all experiments reported, except that of Buss (1953), designs were used that controlled for fortuitous intermitten reinforcements effects in a nonreversal shift.

The mediational hypothesis has generated confusion. Perhaps the following brief statements will clarify some possible areas of misunderstanding.

1. The mediational hypothesis is neither new nor revolutionary. Meyer (1911) and Watson (1913) referred to it, and Hull (1930) gave it a more formal status by coining the concept of the "pure stimulus act." Guthrie (1952) has always laid heavy stress on a mediational-type hypothesis when emphasizing the importance of proprioceptive stimulation in learning.

2. The implicit stimulus and response events *need not* be conceived as having an existence independent of their relation to independent and dependent variables. These implicit events are theoretical constructs. Their epistemological status is closer to such concepts as drive and habit than to directly observable stimulus and response events.

Some mediating events can conceivably and probably will be coordinated to introspective reports, language behavior, muscular movements, and other observable events. Coordinations of this sort can be useful in developing mediational theory. But such coordinations are not *essential* to mediational theory. The fact that genes are not directly observable (at least according to the geneticists consulted) does not interfere with their theoretical and practical usefulness. Even if it were possible to observe a gene directly, it would be necessary to distinguish between it as an observable entity and as a concept within a nomological network. It would be unwise, and strategically shortsighted, to *identify* mediational events with introspective reports or language behavior, or other observable events. The "validity" of the mediational mechanism does not depend on being coordinated with observable events, but depends instead on being utilized in a successful explanatory system.

Figure 4 characterizes reversal and nonreversal shifts in terms of both a single unit S-R analysis and mediational one.[2] It would be predicted, according to a single-unit hypothesis, that if fortuitous intermittent reinforcements were eliminated from a nonreversal shift, it would occur more rapidly than a reversal shift. The reason for this is that at the time of the shift the difference between the strength of the dominant incorrect habit and the to-be-correct habit is much greater for the reversal, as compared to the nonreversal shift. Consequently more training will be required to make the correct habit dominant in a reversal shift. According to the mediational theory the situation is entirely different. A reversal shift enables the subject to utilize the same mediated response. Only the overt response has to be changed. A nonreversal shift, on the other hand, required the acquisition of a *new* mediated response, the cues of which have to be attached to a *new* overt response. Because the old mediational sequence has to be discarded and a new one formed, the nonreversal shift should be executed more slowly than a reversal shift.[3] Thus, if it were possible to eliminate fortuitous intermittent

---

[2] Figure 4 highlights the problem of what are the effective stimuli that are associated to the overt response in both a reversal and nonreversal shift. It is not intended to be a detailed analysis of which there may be several alternatives. For example, in a single-unit theory the habit to choose the large container might result from learning two separate specific habits (e.g., the choice of a large black container when coupled with a small white one and the selection of a large white container when paired with a small black one). Another possibility, which would be consistent with Spence's theory (1936), is that the response is to the effective stimulus *large* since responses to the other features of the environment are not consistently reinforced. Similarly adult subjects in a reversal shift might use the mediator *size* or *large* or both. The effective stimulus which is controlling the organism's response must be determined by experimentation. The point made here is that the general implications of the single-unit and mediational theories, as discussed in this paper, would be the same for a number of different effective stimuli.

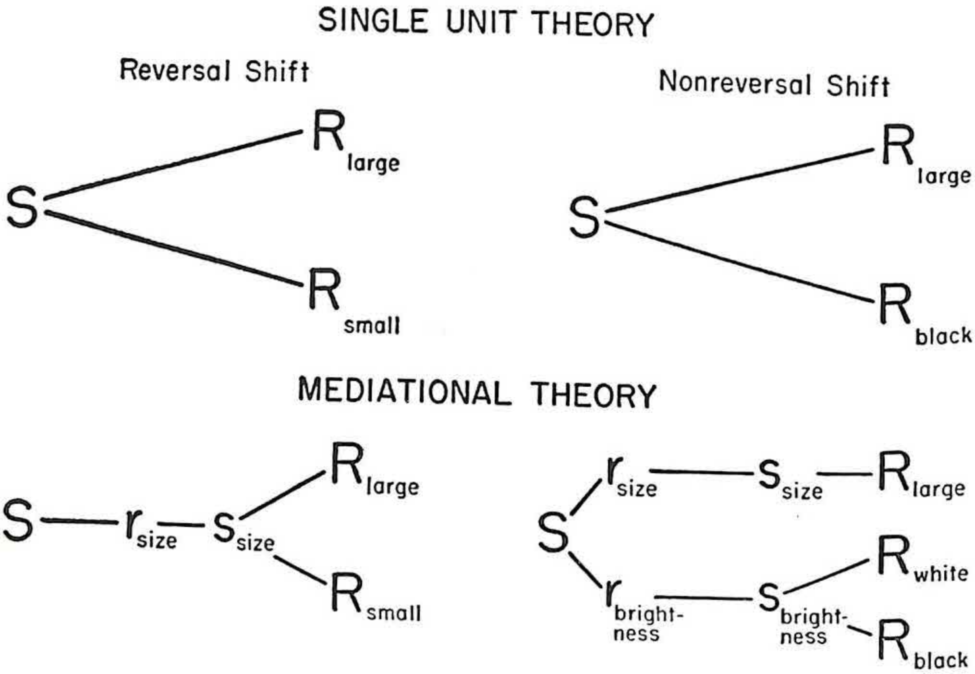


FIG. 4. *A single unit and mediational S-R analysis of a reversal and nonreversal shift.*

---

[3] There are two possible ways of analyzing the superiority of a reversal shift over a nonreversal shift within an S–R mediational framework. One is to simply count the number of new associations that have to be formed. As Figure 4 indicates only one new association has to be formed in a reversal shift while two have to be formed for a nonreversal shift. Another possibility is that a mediating response is more difficult to extinguish than is an overt response. For the present the formulation can remain open-ended until information relevant to these two alternatives is gathered.

reinforcements, then the stage would be set for a crucial experiment testing the conflicting implications of the single-unit and mediational S-R theories. The results of a series of such crucial experiments (Buss, 1956; Harrow & Friedman, 1958; Kendler & D'Amato, 1955) have been consistent with the mediational formulation in showing that college students execute a reversal shift more rapidly than a nonreversal shift. It is important to note that in a similar kind of problem rats find a nonreversal shift easier than a reversal shift (Kelleher, 1956). Thus, one is forced to conclude that a single-unit S-R theory accurately represents the behavior of rats, while mediational S-R theory is required for the concept learning of articulate humans.

Table 1. Mean Number of Trials to Criterion on Test Discrimination for Subjects Scoring above and below the Median on the Training Discrimination

| | Performance on Training Discrimination | |
|---|---|---|
| Group | Above Median (slow learners) | Below Median (fast learners) |
| Reversal | 24.4 | 6.0 |
| Nonreversal | 9.0 | 15.5 |

The discontinuity between the behavior of rats and college students directs one's attention toward the conditions responsible for the development of mediational processes. Somewhere on a hypothetical evolutionary dimension between the rat and college student there should be a point where a transition is made from a single-unit to mediational control. An obvious place to locate this point would be in the behavior of young children.

A study with kindergarten children (Kendler & Kendler, 1959) showed that these children as a group executed a reversal and nonreversal shift at approximately the same rate. One might conclude that the point in human development was discovered which was psychologically halfway between the white rat and the college student, since the kindergarten children were neither responding in a single-unit nor mediational manner, but instead in some compromise fashion. Another possibility is that the children had reached a transitional stage in development, in which the task to which they were subjected led some to function on a single-unit basis, and others to operate with a mediational mechanism. If half of the subjects respond in each way, the total results would have revealed no difference between the two kinds of shifts.

The second alternative seems to fit the data. When the kindergarten children were divided into fast and slow learners on the basis of their performance in the first problem (training discrimination), slow learners performed during the second problem (test discrimination) according to the single-unit theory; like rats they found a nonreversal shift easier. Fast learners, on the other hand, performed in accordance with the mediational theory; like college students, they found a reversal shift easier. These results were interpreted as demonstrating that these kindergartners, taken as a group, were in the process of developing mediating responses relevant to this task, and that some were further along than others.

If this interpretation be correct, then it would follow that for a group of younger (i.e., preschool) children a still smaller proportion should develop appropriate mediating responses. It would be expected that such a group, taken as a whole, would show clearcut evidence of the superiority of a nonreversal over a reversal shift. An experiment (Kendler, Kendler, & Wells, 1960) designed to test this hypothesis produced results consistent with this prediction; like rats, nursery school children found a nonreversal shift to be easier than a reversal shift.

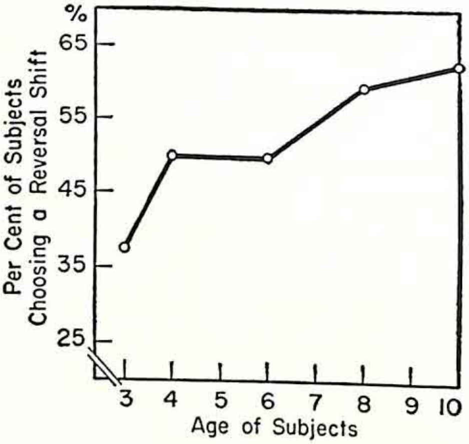


FIG. 5. *Percentage of children responding in a reversal shift manner as a function of age.*

In a very recent study the experimental procedure was modified so that after learning the initial discrimination, the children of 3, 4, 6, 8, and 10 years of age who served as subjects, had a choice of either responding in a reversal or a nonreversal manner. Under such circumstances, it would be expected that the proportion of children who respond in a reversal manner would increase with age. Figure 5 shows that the percentage of children who chose a reversal shift rose gradually from 37.5 at 3 to 62.5 at 10.

Generalizing from all of these results, it would seem that in their early development, children tend to respond in a manner consistent with a single-unit S-R theory. With age, they develop a tendency to respond in a mediational manner. The last study cited suggests that it is, or will soon be, possible to ascertain the lawful relationship governing the course of this development.

The point of these experiments is not to classify children into one of two categories: rat-like or human-like. Their aim is to lay the groundwork for experiments designed to investigate the mediational process itself. If one wants to investigate the mediational processes, does it not seem sensible to scrutinize them at the time when they are developing? Answering this question in the affirmative, it was decided to investigate the relationship between the hypothesized mediational processes and verbal behavior—a relationship everybody assumes to be intimate and important.

Particularly relevant to this attempt to coordinate verbalization with mediation were observations that during the course of the experiments just described, it was not uncommon for children to verbalize spontaneously the correct solution while simultaneously making an incorrect choice. A few children did this for many consecutive trials. This observation is relevant to the concept of vertical processes. Two chains of habits are occurring simultaneously. One has to do with verbal response; the other with the overt choice. For these children the two chains are parallel, that is, they do not interact.

Luria (1957), the Russian psychologist, made somewhat similar observations in his research with children. He explains this sort of phenomenon in the following way:

> In the early stages of child development, speech is only a means of communication with adults and other children. . . . Subsequently it becomes also a means whereby he organizes his own experience and regulates his own actions. So the child's activity is mediated through words (p. 116).

These observations and their interpretations of noninteracting parallel processes point to the complex interrelationships existing between verbal behavior on the one hand and problem solving on the other. If nothing else, they destroy the illusion that it is reasonable to describe an organism as verbal or nonverbal without considering the problem with which it is confronted. The terms verbal and nonverbal become meaningful—and fruitful—when related to specific problem solving tasks.

It would seem fruitful to investigate the cue function of words for children of two age levels. One possibility is that age influences problem solving only in so far as it leads to the acquisition of words. If younger children, say 4 years of age, could acquire the same words as 7-year-olds, they would solve a simple concept learning problem the same way. The other possibility is that the acquisition of the verbal label by itself is not sufficient; the word must be integrated with other behavioral chains to influence problem solving behavior. And for this to happen some developmental changes must first take place.

In order to test these two alternatives, children of 4 and 7 years of age were presented with another variation of the reversal shift problem as shown in Figure 6. They initially learned a simple discrimination between a pair of stimuli that varied simultaneously in size and brightness. In the illustration provided in Figure 6, the large black square is correct. While they were learning, the children were required to verbalize aloud the stimuli to which they were responding. One-third learned to say "large" (or "small" as the case may be) by the simple device of instructing them to tell the experimenter which was correct, the large or the small one. Another third learned to say "black" (or "white") in a corresponding way. The remaining third was not required to say anything. After learning the discrimination, all subjects were presented with a reversal shift. In the example depicted in Figure 6, the shift is to small regardless of size. Thus, the group that initially described the correct stimulus as "large" had verbalized the relevant dimension. The verbal response of "black" was irrelevant to this reversal shift.

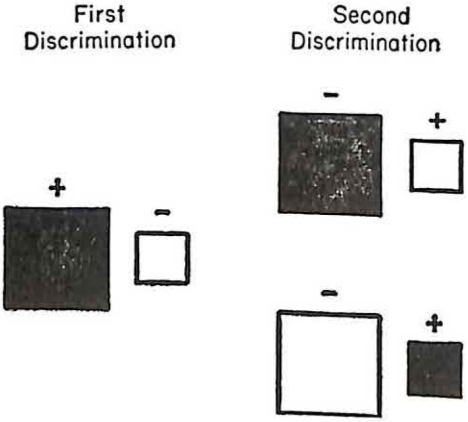


FIG. 6. *The experimental procedure used to study the influence of verbal habits on a reversal shift.*

Figure 7 shows the results of the three experimental groups for the two age levels. If developmental processes affect the utilization of verbal responses in problem solving, then it would be expected that the three verbalization conditions (which produced a significant main effect) would influence the behavior of the two age groups differently. These results suggest, but not quite at a significant level, that there is an interaction effect. Figure 7 shows that the younger children profited by making the kind of verbal response appropriate to a reversal shift, while they were hindered by learning inappropriate verbal responses. With no verbalization the 7-year-old children who presumably were responding largely in a mediational manner, accomplished a reversal shift much more rapidly than their younger counterparts. But unlike the 4-year-olds, they did not profit from being trained to make the relevant responses. At 7 years of age they are capable of making the response themselves and outside help appears to be of little use. In contrast, the influence of irrelevant verbalizations is marked. The performance of the 7-year-olds was even poorer than that of the 4-year-olds, suggesting that the interfering effects of being given an inappropriate mediated response are greater when one is capable of spontaneously generating the correct one (7-year-olds) than when one is not (4-year-olds).

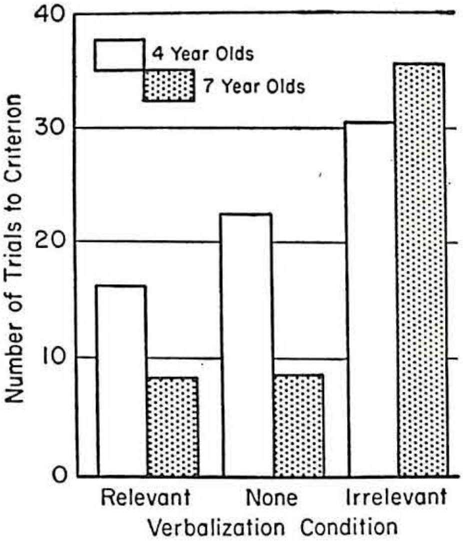


FIG. 7. *The effect of verbalizations on a reversal shift for 4- and 7-year-old children.*

How are these data to be explained? Attributing differences to developmental factors is not sufficient. It is necessary to represent development differences in terms of the concepts of the behavior model that is being used. That is, if a verbal label for a young child does not possess the same cue function as it does for an older child, then it becomes necessary to specify how and why this comes about. To some extent this has been done by emphasizing the transition from a single-unit to a mediational system, as well as suggesting that with age an increase occurs in interaction among chains of different vertical levels. But obviously this analysis of the developmental process demands further theoretical and empirical development.

These studies are intimately related to the oft-reported finding that many species of subhuman animals are able to make a fairly rapid reversal shift *if* they receive a previous series of such shifts. Rats (Buytendijk, 1930; Krechevsky, 1932) show a marked improvement in executing successive reversals. They finally reach a point (Dufort, Guttman, & Kimble, 1954), in a **T** maze, where they learn to go to a new rewarded goal after making only one error. Even more dramatic are the rapid discrimination reversals exhibited by Harlow's (1949) monkeys. But fish (Wodinsky & Bitterman, 1957) exhibit only a slight improvement in successive reversals, while isopods (invertebrates) show no improvement (Thompson, 1957).

Because of the necessity to use somewhat different experimental procedures for different species, it is difficult to draw an unqualified conclusion about the ability of different species to transfer what has been learned from previous reversal shifts to a new one. But the suggestion is strong that as you ascend the evolutionary scale organisms acquire a greater capacity to generate cues that enable them to make rapid reversal shifts. This behavior, according to our analysis, borders on the language responses of humans. The main difference is that our human subjects, except those of a very young age, exhibit rapid reversals without any previous reversal training. Whereas the human automatically seems to generate a mediated response that provides the basis for his rapid reversal, the animal subject must gradually acquire an ability to respond appropriately to some response-produced cue resulting from nonreinforcement of a previously correct response.

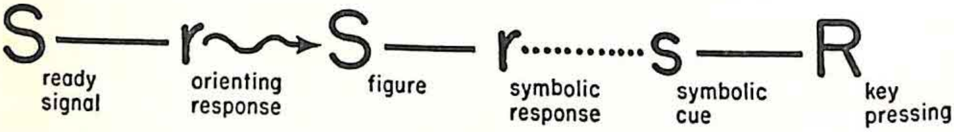


FIG. 8. *The hypothesized behavioral chain operating at the time the subject was being shifted to the second concept. (Capital letters refer to directly observable stimulus and response events, while small letters refer to those that are inferred.)*

Up to now, the reversal and nonreversal technique has been used to investigate mediational and developmental variables. It has proved sufficiently flexible to be used in a study (Kendler, Glucksberg, & Keston, 1961) which was designed to lengthen a problem solving chain so that the interaction between various segments could be observed. In this study a perceptual orienting S-R unit was added on to the mediational chain already described. Figure 8 illustrates in an oversimplified manner the behavioral sequence involved in this study in which subjects had to learn to press the correct button when two physically discrete and spatially separate stimulus patterns were projected on a screen at such a rapid rate that only one could be perceived on any trial. During the learning of each of two successive concepts (involving either a reversal or nonreversal shift), the subject had to pay attention to the relevant stimulus pattern while ignoring the irrelevant one. Thus, in order to make the correct overt response consistently, a subject initially had to make the appropriate orienting response in order to perceive the relevant stimulus pattern to which he had to make the correct mediational response which served as the cue for the key-pressing act.

An experimental design was used in which, at the time of the shift from the first to the second concept, one group had already learned the appropriate orienting response as well as the appropriate mediating act. They needed only to learn a new terminal key-pressing response. The shift, for them, was easy to make. In contrast, the behavior of three other experimental groups was significantly worse. One group had to learn a new orienting response, e.g., look to the left instead of the right. Another group had to learn a new mediated response (i.e., they were required to make a nonreversal shift). The last group had to acquire both a new orienting and mediated response. The fact that the groups which were missing one or both of the necessary behavior units (orienting and mediated responses) did not differ significantly among themselves, as well as being much poorer than the group that had both, highlights the problem of synchronizing the S-R units in a behavioral chain. The advantage in this study of having one appropriate unit without the other is at best negligible. The reason for this is that reinforcement is only achieved consistently when both the appropriate orienting and mediating responses are operating. This particular study points to the need for discovering laws associated with the strengthening and weakening of independent S-R units in a problem solving chain, as well as the principles governing their synchronization.

This study also highlighted a very basic problem in all of these reversal studies. This problem has to do with the very first correct response following the reversal shift. After discovering that the previous mode of responding is erroneous, what makes the subject change his response, i.e., push the button that was previously wrong? Introspective reports fail to provide any clearcut answer and even if they did they would be in need of explaining (Kendler, 1961).

One hypothesis is that the selection of the new correct response is due to the operation of a behavioral chain in addition to the one described in Figure 8. The first nonreinforcement in a reversal shift sets off a chain, the consequence of which is to select the response other than the one that was previously correct. This may result from a number of different reasons (e.g., logical considerations, forgetting, etc.). The important point, however, is that the new key-sorting response occurs contiguously with the implicit mediational response appropriate to a reversal shift. As a result, a new association is formed between the old implicit cue and the new key-pressing response.

In essence, what is being stated is that adult subjects, when making or deciding to make the first correct postshift response, do not adopt the *principle* underlying a reversal shift. Instead, it is assumed processes are operating which encourage the selection of the correct response while an implicit cue appropriate to a reversal shift is operating. This sort of an analysis was described previously (Kendler & Mayzner, 1956) as

> sort of a James-Lange theory of problem solving . . . one makes the overt correct . . . response and if the appropriate symbolic cue is present, then problem solution will occur (pp. 247–248).

Guthrie (1959) says the same thing more neatly: "*What is being noticed becomes a signal for what is being done*" (p. 186).

Again the authors would like to guard against giving the impression of oversimplifying a terribly complex problem. They do not believe the contiguous occurrence of an implicit cue from one chain with the correct overt response from another tells the whole story. This new association in order to persist must be reinforced and in some manner "fit into" the subject's ongoing behavioral chains.

The emphasis on this vertical connection between a cue and a response from different chains is related in a distant way to Hebb's (1958) stressing the role of "chance" in problem solving:

> There are few scientists who have not had the experience of setting out to solve problem A and ending up instead with the solution to B. . . . This is serendipity, the art of finding one thing while looking for another (p. 215).

According to the present analysis, serendipity results from the adventitious and contiguous occurrence of a cue and a response which are themselves segments from different behavior chains. Theoretically it should be possible to demonstrate this point experimentally by training subjects to respond simultaneously to two separate tasks. A problem then would have to be presented that requires for its solution the combination of a stimulus from one chain with the response from the other. In such an experimental situation, controlling the time relationship between the two should have an important effect on problem solving. Presumably contiguity between the two should provide the most optimal conditions for problem solving (Underwood, 1952). The development of this kind of experimental procedure should allow for parametric studies of the basic variables of the phenomenon which has commonly been called "insight," as well as throw light upon issues raised by others (e.g., Cofer, 1957; Maltzman, 1955; Saugstad, 1957).

The pretheoretical model that guides the present research has many more facets that can be exploited. Only one will now be mentioned. Glucksberg

(1962), for example, extended neobehavioristic drive theory (Spence, 1956) to problem solving. He used a functional-fixedness problem (Adamson, 1952, Duncker, 1945) in which the correct response in the habit hierarchy could either be made to be low or high. If the correct habit was low, it would be expected that a strong drive would retard problem solving because it would retard the extinction of the dominant incorrect response (Kendler & Lachman, 1958; Perin, 1942). Since drive energizes behavior, a high drive should facilitate problem solving performance when the correct habit is dominant. The findings were consistent with this analysis.

Because functional-fixedness problems are often represented in perceptual terms, Glucksberg was interested in seeing whether the same drive model could be applied to a simple perceptual recognition problem in which subjects were instructed to identify tachistoscopically presented words as rapidly as possible. The results were similar to those reported for the functional-fixedness study: when the correct response was dominant, an increase in drive improved performance, i.e., the visual duration threshold was lowered. In contrast, increasing drive when the correct response was low in the hierarchy raised the threshold.

There is obviously still much more work, both empirical and theoretical, needed to develop the model that has been described. At this point it may be appropriate to summarize the major points of this paper.

There is not just one way to investigate problem solving. The researcher who is interested in problem solving has several different pretheoretical models from which to choose. This paper reported the results of a research program based on an S-R model in which the importance of horizontal and vertical processes were emphasized. Horizontal processes refer to the linking of successive S-R units into a behavorial chain, while vertical processes refer to the assumption that independent chains occur simultaneously. A series of experiments was reported, the implications of which supported postulating a mediational mechanism within a behavioral chain. By comparing the behavior of human subjects of different ages, as well as relating their results to lower animals, it was possible to infer that as a child matures he makes a transition from responding on the basis of a single-unit S-R mechanism to a mediational one. Additional data were cited that suggest the full impact of verbal behavior on problem solving depends on developmental processes that encourage interaction between chains at different vertical levels. It was also suggested that problem solving begins in a simple concept learning task when a correct overt response from one behavioral chain occurs contiguously and adventitiously with the appropriate implicit cue from another chain. The paper was concluded by citing findings that suggested the neobehavioristic drive theory which assumes that the effect of different levels of drive depends on the position of the correct response in the habit hierarchy is applicable to a functional-fixedness problem as well as a perceptual-recognition task.

If nothing else, it is hoped that the present paper demonstrates that it is possible to investigate problem solving in a systematic fashion. If more psychologists accepted this possibility and were willing to expend their research energies in the field of problem solving, progress in this area would be greater than it is today.

## REFERENCES

Adamson, R. E. Functional-fixedness as related to problem solving. *J. exp. Psychol.*, 1952, 44, 288–291.

Buss, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discrimination. *J. exp. Psychol.*, 1953, 45, 75–81.

Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, 52, 162–166.

Buytendijk, F. J. J. Über das Umlernen. *Arch. Néerl. Physiol.*, 1930, 15, 283–310.

Cofer, C. N. Reasoning as an associative process: III. The role of verbal responses in problem solving. *J. gen. Psychol.*, 1957, 57, 55–58.

Dufort, R. H., Guttman, N., & Kimble, G. A. One trial discrimination reversal in the white rat. *J. comp. physiol. Psychol.*, 1954, 47, 248–249.

Duncan, C. P. Recent research on human problem solving. *Psychol. Bull.*, 1959, 56, 397–429.

Duncker, K. On problem solving. *Psychol. Monog.*, 1945, 58, No. 5 (Whole No. 270).

Glucksberg, S. The influence of strength of drive on functional fixedness and perceptual recognition. *J. exp. Psychol.*, 1962, in press.

Guthrie, E. R. *The psychology of learning.* (Rev. ed.) New York: Harper, 1952.

Guthrie, E. R. Association by contiguity. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 158–195.

Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, 56, 51–65.

Harrow, M., & Friedman, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, 55, 592–597.

Hebb, D. O. *A textbook of psychology.* Philadelphia: Saunders, 1958.

Hull, C. L. Knowledge and purpose as habit mechanisms. *Psychol. Rev.*, 1930, 37, 511–525.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Kelleher, R. T. Discrimination learning as a function of reversal and nonreversal shifts. *J. exp. Psychol.*, 1956, 51, 379–384.

Kendler, H. H. Problems in problem solving research. In, *Current trends in psychological theory: A bicentennial program.* Pittsburgh: Univer. Pittsburgh Press, 1961.

Kendler, H. H., & D'Amato, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, 49, 165–174.

Kendler, H. H., Glucksberg, S., & Keston, R. Perception and mediation in concept learning. *J. exp. Psychol.*, 1961, **61**, 186–191.

Kendler, H. H., & Kendler, T. S. Inferential behavior in preschool children. *J. exp. Psychol.*, 1956, **51**, 311–314.

Kendler, H. H., Kendler, T. S., Pliskoff, S. S., & D'Amato, M. F. Inferential behavior in children: I. The influence of reinforcement and incentive motivation. *J. exp. Psychol.*, 1958, **55**, 207–212.

Kendler, H. H., & Lachman, R. Habit reversal as a function of schedule of reinforcement and drive strength. *J. exp. Psychol.*, 1958, **55**, 584–591.

Kendler, H. H., & Mayzner, M. S., Jr. Reversal and nonreversal shifts in card sorting tests with two or four sorting categories. *J. exp. Psychol.*, 1956, **51**, 244–248.

Kendler, H. H., & Vineberg, R. The acquisition of compound concepts as a function of previous training. *J. exp. Psychol.*, 1954, **48**, 252–258.

Kendler, T. S., & Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58**, 56–60.

Kendler, T. S., & Kendler, H. H. Inferential behavior in children: II. The influence of order of presentation. *J. exp. Psychol.*, 1961, **61**, 442–448.

Kendler, T. S., Kendler, H. H., & Wells, D. Reversal and nonreversal shifts in nursery school children. *J. comp. physiol. Psychol.*, 1960, **53**, 83–88.

Koch, S. *Psychology: A study of a science.* Vol. 1. *Sensory, perceptual, and physiological formulations.* New York: McGraw-Hill, 1959.

Kohler, W. *The mentality of apes.* New York: Harcourt, Brace, 1925.

Krechevsky, I. Antagonistic visual discrimination habits in the white rat. *J. comp. Psychol.*, 1932, **14**, 263–277.

Lachman, R. The model in theory construction. *Psychol. Rev.*, 1960, **67**, 113–129.

Luria, A. R. The role of language in the formation of temporary connections. In B. Simon (Ed.), *Psychology in the Soviet Union.* Stanford: Stanford Univer. Press, 1957. Pp. 115–129.

Maier, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, **10**, 115–143.

Maltzman, I. Thinking: From a behavioristic point of view. *Psychol. Rev.*, 1955, **62**, 275–286.

Meyer, M. F. *The fundamental laws of human behavior.* Boston: Gorham, 1911.

Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.

Saugstad, P. An analysis of Maier's pendulum problem. *J. exp. Psychol.*, 1957, **54**, 168–179.

Skinner, B. F. *Science and human behavior.* New York: Macmillan, 1953.

Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Thompson, R. Successive reversal of a position habit in an invertebrate. *Science*, 1957, **126**, 163–164.

Toulmin, S. *The philosophy of science.* London: Hutchinson Univer. Library, 1953.

Underwood, B. J. An orientation for research on thinking. *Psychol. Rev.*, 1952, **59**, 209–220.

Watson, J. B. Psychology as the behaviorist sees it. *Psychol. Rev.*, 1913, **20**, 158–177.

Wodinsky, J., & Bitterman, M. E. Discrimination-reversal in the fish. *Amer. J. Psychol.*, 1957, **70**, 569–575.

# 22

# Behaviors During Problem Solving Among Children of Low, Average, and High Intelligence

## — *Herbert J. Klausmeier and Leo J. Loughlin*

*Considerable research has shown that among school children there are great differences in intellectual and psychomotor abilities as related to learning. Yet the factors underlying the observed differences have been less clearly defined. Here the authors report an investigation of qualitative differences in observed behavior among groups of children of low, average, and high IQ. The children were observed while in the actual process of solving arithmetic problems, and their behaviors were classified in terms of efficiency of method, persistence, and mode of attack. Producing possible solutions, verifying each proposed solution, and applying information from one trial to the subsequent one were the differences of most practical significance noted among the three groups.*

THE RANGE of differences among school children in many intellectual and psychomotor abilities and other characteristics is defined clearly. Less clearly defined are the factors which contribute to the differences. Previously, no statistically significant difference was found among groups of low, average, and high IQ children in retention of arithmetic problems, graded to each child's achievement level (Klausmeier & Feldhusen, 1959); nor in emotional adjustment, integration of self-concept, and achievement

Reprinted from *Journal of Educational Psychology,* 1961, **52,** 148–152, by permission of the senior author and the American Psychological Association.

in relation to expectancy (Klausmeier, Check, & Feldhusen, 1960). Large differences with little overlapping between IQ levels were found in achievements in reading, language, and arithmetic. The purpose of this investigation was to ascertain possible differences among the three groups in efficiency of method, persistence, and mode of attack while in the actual process of solving arithmetic problems.

## Procedure

### Subjects

The subjects were 20 boys and 20 girls with WISC IQs of 55–80, 20 boys and 20 girls with WISC IQs of 90–110, and 20 boys and 20 girls with WISC IQs of 120–146. The 80 children of high and average IQ were enrolled in 15 regular fifth grade classes of two large public school systems; and the 40 low IQ children were enrolled in 14 special classes for educable mentally retarded children in the same schools. Excluded from the study were low IQ children exhibiting a second handicap to a marked degree or definite organic symptoms of retardation such as mongolism. All had birthdates between September 15, 1947, and December 15, 1948; and the mean age at the time of gathering the data was 131 months for each IQ group.

### Methods

The first task was to get problems properly graded in difficulty. After extensive experimenting with pupils in regular classrooms and with low IQ pupils not in the study, a group of 29 problems at three main levels of difficulty was arranged. The problems for the low IQ children were subdivided further with the lowest level intended for those who did not know the number of cents in a dime or who could not perform simple addition. A reserve level of problems of high difficulty was established for the high IQ child who might solve an assigned problem in less than 3 minutes. The least difficult problem designed for a low IQ child was to make 7¢ with three coins and the most difficult designed for a high IQ child was to make $14.53 with 16 bills and coins.

Besides this initial formulation of problems, two other measures were used to make certain that the problem was of appropriate difficulty. A survey test was administered to each child to ascertain his knowledge of our money system, related vocabulary, and mathematical skills. Also, if a child completed his assigned problem in less than 3 minutes, the next problem in order of difficulty was assigned.

Each child solved only one problem. This problem and the survey test mentioned previously were administered by a carefully trained examiner; and another trained person observed the entire procedure. From the time the problem and instructions were given the child until the problem was solved, time in seconds and the incidence of the following behaviors of the child were recorded: (*a*) manipulations of coins by the low IQ children or manipulation of figures with pencil and paper by the average and high IQ children; (*b*) engaging in nontask activities or withdrawal which resulted in the examiner's encouraging the child to continue his problem solving; (*c*) offering an incorrect solution; (*d*) noting and correcting mistakes independently; (*e*) verifying the solution by checking the amount of money independently; (*f*) verifying the solution by checking the number of coins independently; and (*g*) using a random, logical, or a combination approach to the solution. In random approach the subject merely manipulated or explored the materials without an apparent plan of attack. In logical approach the subject made deliberate manipulations to test successive trials or hypotheses. Some children changed from an apparently random to a logical approach, and vice versa.

Efficiency of the child's method was defined as the ratio of moves (coins or paper-pencil computations made) to the least number needed to achieve the solution. For example, the least efficient subject made 39 moves when only 2 were required, with a resulting efficiency ratio of .051. Any instance of a child's being encouraged by the examiner to cease from a nontask activity or to continue at the task was recorded as the measure of nonpersistence. The mode of attack was observed in the verbal responses of the subject and in his manipulations of coins or figures, as described in the preceding paragraph, *c* through *f*.

The data secured on each subject were tabulated and combined according to the IQ groups. Analysis of variance was used to compare efficiency of method and chi square to compare incidences of the other behaviors among the three groups.

## Results

The extent to which the problems were of equal difficulty for each IQ group was ascertained by comparing the time in seconds required to reach a solution to the problem. The low IQ group required a mean time of 624.17 seconds, the average group 627.65, and the high group 578.47 seconds. On the basis of a small and insignificant $F$ value obtained among the means, it was concluded that the assigned problems were about equally difficult for the three IQ groups.

A summary of the data and the results of the tests of statistical significance of differences among the three IQ groups and between groups are presented in Table 1. Column 1 lists the observed behaviors. Columns 2–4 contain the scores or number of subjects in each IQ group who demonstrated the

TABLE 1. Summary of Observed Scores and Results of Tests of Statistical Significance of Differences

| | Score or Incidence | | | Significance of Differences | | | |
|---|---|---|---|---|---|---|---|
| | | | | Among Groups | Between Groups | | |
| Behavior | High | Average | Low | | Hi-Av | Hi-Low | Av-Low |
| Efficiency of Method (%) | .353 | .309 | .237 | $.05>p>.01$ | *ns*° | $.05>p>.01$ | *ns* |
| Nonpersistence (Incidence) | 3 | 10 | 17 | $.01>p>.001$ | $.02>p>.01$ | $.01>p>.001$ | $.02>p>.01$ |
| Mode of Attack (Incidence) | | | | | | | |
| Offered Incorrect Solution | 10 | 16 | 33 | $.001>p$ | $.001>p$ | $.001>p$ | $.01>p>.001$ |
| Noted and Corrected Mistakes | 14 | 12 | 5 | *ns* | | | |
| Verified Solution—Amt. Money | 38 | 31 | 11 | $.001>p$ | $.001>p$ | $.001>p$ | $.001>p$ |
| Verified Solution—No. Coins | 36 | 26 | 9 | $.001>p$ | $.001>p$ | $.001>p$ | $.001>p$ |
| Random Approach | 1 | 8 | 21 | $.001>p$ | $.01>p>.001$ | $.001>p$ | $.001>p$ |
| Logical Approach | 14 | 5 | 3 | $.01>p>.001$ | $.01>p>.001$ | $.01>p>.001$ | $.05>p>.02$ |

° Not significant at .05 level.

behavior. Column 5 shows the results of the tests of statistical significance of differences among the three IQ groups. Columns 6–8 give the results of the tests of significance of differences between groups—high and average, high and low, and average and low.

Table 1 shows that the low IQ group was inferior to the high IQ and to the average IQ group in seven and six of the observed behaviors, respectively; the average IQ group was inferior to the high IQ in six behaviors. Although the mean number of seconds to solve was nearly identical for the three groups, the incidences of behaviors were greatly different. To clarify the differences, short sketches of the procedures used by a high and a low IQ boy are now given.

Richard had a WISC IQ of 138 and used 585 seconds to solve the problem: "Make $9.77, using 12 bills and coins." He studied the situation and during the first 3 minutes made a first attempt at solution, getting the amount correct but using 13 instead of 12 bills and coins—9 one dollar bills, 1 half-dollar, 1 quarter, and 2 pennies. He checked the amount and number of coins and started over. During the next 3 minutes he made his second and third tries, each time starting with a five dollar bill. On the second he had 12 bills and coins but an incorrect amount; on the third try, neither was correct. Each time he checked amount and number of coins. During the third 3 minutes he made his fourth try, but made a mistake in adding the amount. Checking through his solution at the beginning of the fourth 3-minute period, he found the error and after 45 seconds offered as a correct solution 1 five dollar bill, 4 one dollar bills, 1 half dollar, 1 dime, 3 nickels, and 2 pennies. Richard made 25 separate moves, but only 6 were needed, as shown in the final solution. Therefore, his efficiency ratio was 6/25, or .24.

In terms of a problem solving sequence set forth by Merrifield, Guilford, Christensen, and Frick (1960), Richard (*a*) accepted the task as a problem and set a goal of achieving a correct solution, (*b*) analyzed the available information in relation to his goal, (*c*) produced a solution and independently rejected it (verification), (*d*) applied knowledge from this and subsequent solutions to each successive solution produced (reapplication), and (*e*) finally verified and accepted the correct one.

Michael, one of the more able low IQ boys, had a WISC IQ of 73 and used 805 seconds to "Make 46¢, using 4 coins." Ten coins of each denomination—penny, nickel, dime, and quarter—had been placed on the table before him. (In a pretest it had been established that he had the requisite skill in adding and knowledge of coins to solve the problem.) Michael, as Richard, restated the problem correctly prior to attempting a solution. During the first 3 minutes, Michael took 3 dimes and 10 pennies from the supply, then returned to the supply 3 pennies, one at a time, and offered the remainder as a correct solution. Upon being told that it was incorrect and with all coins returned to the supply, he studied the coins and offered 1 quarter and 2 dimes as a second solution during the second 3 minute period. During the third 3 minutes, he withdrew from the supply 1 quarter, 1 dime, 1 nickel, and 1 penny and offered this as correct. When told it was incorrect and the coins had been returned, he again studied the coins, and during the fourth 3 minutes withdrew 1 quarter, 1 penny, 1 dime, then put the quarter back. During the fifth 3 minutes he again withdrew the quarter and another dime from the supply, thus having 1 quarter, 2 dimes, and 1 penny. This he offered as correct without observably checking accuracy of amount or number of coins. He made a total of 16 coin movements when only 3 were needed; his efficiency ratio was .188. At no point was it necessary to encourage Michael to continue or to keep at the task; however, like the majority of low IQ children, he offered incorrect solutions and did not verify amount of money nor the number of coins.

In what respects are the low IQ most different from the high IQ children? As noted in the above descriptions, the high IQ boy was markedly superior in verification and reapplication, eventually arriving at the correct solution. Further, other low IQ children, different from Michael, did not persist when informed that their solutions were wrong, and after two or three such experiences, manipulated the coins without producing a close approximation of the final correct solution.

## Discussion

If time to reach a correction solution had been used as the criterion of efficiency of the problem solving, no significant difference would have been observed among the three groups. Using the ratio of efficiency as the criterion, a significant difference was observed. The differences of most practical significance among the groups were connected with producing possible solutions, verifying each proposed solution, and applying information from each preceding trial to the subsequent one. As John (1957), Duncan (1959), and others have pointed out, processes such as these used in arriving at solutions are the important variables in understanding efficiency of performance and learning.

Whether more of the low IQ children in this study could learn seemingly simple processes as deciding a plan for making 13¢ with four coins is not known. Similarly, verifying the amount of money and number of coins does not appear to be difficult, although the high IQ children also tended to do only one or the other on the first try. If these processes can be learned, more specific instruction in school devoted to helping children acquire processes, rather than solutions, would be helpful.

Though not studied intensively, differences within each IQ group were large. For example, the ratio of efficiency varied from .10 to .80 among the high IQ children. Further, 10 high and 16 average IQ children offered incorrect solutions as shown in Table 1. Help given to children of average and high abilities when needed might be especially profitable in achieving improved problem solving performances.

## Summary

Differences among groups of children of low, average, and high IQ were found in observed behaviors as they solved problems, graded in difficulty to their present achievement levels. The high IQ children showed a greater incidence than the average and the low, and the average a greater incidence than the low in noting and correcting mistakes independently, verifying solutions, and using a logical approach; whereas the low IQ showed a greater incidence than the average and the high, and the average a greater incidence than the high in nonpersistence, offering an incorrect solution, and using a random approach. The high IQ children were superior to the low in efficiency of method, as ascertained by the ratio of number of moves made to the least number required for solution of the problem. Differences in performances among individuals within IQ groups were also large.

## REFERENCES

Duncan, C. P. Recent research on human problem solving. *Psychol. Bull.*, 1959, **56**, 397–429.

John, E. R. Contributions to the study of the problem-solving process. *Psychol. Monogr.*, 1957, **71**, No. 18 (Whole No. 447).

Klausmeier, H. J., Check, J., & Feldhusen, J. Relationships among physical, mental, achievement, and personality measures in children of low, average, and high intelligence at 125 months of age. *Amer. J. ment. Defic.*, 1960, **65**, 69–78.

Klausmeier, H. J., & Feldhusen, J. Retention in arithmetic among children of low, average, and high intelligence at 117 months of age. *J. educ. Psychol.*, 1959, **50**, 88–92.

Merrifield, P. R., Guilford, J. P., Christensen, P. R., & Frick, J. W. A factor-analytic study of problem-solving abilities. *Rep. Psychol. Lab., Univer. Southern Calif.*, 1960, No. 22.

# 23

# A Definition of Critical Thinking[1]

## — *Robert H. Ennis*

*In this article, critical thinking is conceived as "the correct assessing of statements." The author examines nine major aspects of critical thinking and suggests that these may serve as an "elementary and intermediate statement of goals in the teaching of critical thinking." Although no attempt is made to instruct the reader how to teach critical thinking, the provocative examples presented suggest fruitful teaching procedures.*

AS A ROOT notion, "critical thinking" is taken to mean the correct assessing of statements.[2] Since there are various kinds of statements, various relations between statements and their grounds, and various stages in the

---

[1] This is a streamlined version of my article, A concept of critical thinking, *Harvard educ. Rev.*, 1962, **32** (1), 81–111. Two features of that article are omitted here: the theoretical analysis (using three dimensions) of critical thinking, and the proposals for needed research. The latter feature is expanded in an article entitled Needed: research in critical thinking, *Educ. Leadership*, 1963, **21** (1), 17–20, 39.

The feature of the paper in *Harvard Educational Review* which is retained here is the main part of that paper—the listing of a number of aspects of critical thinking together with a provisional statement of criteria to help in making the judgments called for. The current list is different in two ways: it has been slightly reshuffled and one aspect of critical thinking—judging definitions—has been omitted in order to reduce the length of this version of the paper. A more extended treatment would include this aspect as well as an additional one, judging explanations.

Further refinement of this list is a continuing task, and of course much remains to be done.

---

As indicated in the author's footnote above, this is a modified version of his article "A Concept of Critical Thinking" which appeared in the *Harvard Educational Review*, 1962, **32** (1), 81–111. It is reprinted, in its present form, by permission of the author and the publisher.

[2] This basic notion was suggested by B. Othanel Smith (1953, p. 130): "Now if we set about to find out what . . . [a] statement means and to determine whether to accept

process of assessment, we can expect that there will be various ways of going wrong when one attempts to think critically. In view of this fact, the list of aspects of critical thinking about to be presented, which may be looked upon as a list of specific ways to avoid the pitfalls in assessment, is bound to be a rather heterogeneous list.

This list and the accompanying criteria for judging statements are based in a large part upon a study of the literature in education, philosophy, and psychology. Items that were of help will be found in the reference list at the end of this article. This list of critical thinking aspects is also based upon an analysis of a number of specimens of alleged justifications of statements, and a consequent realization of the places where these justifications can go wrong. You may look upon this list as a statement of a number of items that, if taught to students, will result in a greater likelihood that they will be critical thinkers.

## A List of Major Aspects of Critical Thinking

A critical thinker is characterized by proficiency in judging whether:

1. A statement follows from the premises.
2. Something is an assumption.
3. An observation statement is reliable.
4. A simple generalization is warranted.
5. A hypothesis is warranted.
6. A theory is warranted.
7. An argument depends on an ambiguity.
8. A statement is overvague or overspecific.
9. An alleged authority is reliable.

Although the root notion calls for its inclusion, the judging of value statements is deliberately excluded from the above list. This exclusion admittedly weakens the attractiveness of the list, but makes the job more manageable. So long as we remember that this exclusion has occurred, we should not be confused by the truncated concept. Perhaps this gap can at some future time be at least partially filled.

---

or reject it, we would be engaged in thinking which, for lack of a better term, we shall call critical thinking." Since Smith's definition does not use any words like "correct," his notion is slightly different.

Smith's concept of critical thinking permits us to speak of "good critical thinking" and "poor critical thinking" without redundance or contradiction. Though this is an accepted manner of speaking, the predominant manner of speaking presumably builds the notion of correct thinking into the notion of critical thinking. Though the latter interpretation is used in this paper, it would be easy to restructure what follows and use Smith's concept. "Good critical thinking" in Smith's sense means the same as "critical thinking" as used in this paper.

The exclusion of other important kinds of thinking (creative thinking, for example) from this basic concept of critical thinking does not imply that the others are unimportant, nor does it imply that they are separable from it in practice. This exclusion is simply the result of an attempt to focus attention on one important kind of thinking.

Another item which has deliberately been excluded from the list is proficiency in judging whether a problem has been identified. This is excluded not because it is unimportant, but because it resolves into one or another of the items on the list (or the judging of value statements) in each of the various meanings of "identifying a problem." This point will be developed later.

## Clarification of the Major Aspects of Critical Thinking

In this section each of the listed aspects of critical thinking will be examined, and, if possible, criteria will be presented, clarified, and, when it seems necessary, at least partially justified.

### 1. Judging whether a statement follows from the premises

The concern of most logic books is with whether a statement follows *necessarily* from the premises. This is the judging of deduction. Reasoning in mathematics, "if-then" reasoning, and syllogistic reasoning all exemplify deduction.

The basic criterion is this: "A conclusion follows necessarily, if its denial contradicts the assertion of the premises." Various rules have been developed for different types of deduction, but all see to it that this requirement is fulfilled. Well-developed sets of rules include:

1.1 The rules for handling equations and inequalities.
1.2 The rules of "if-then" reasoning:
  1.21 Denial of the "then-part" requires denial of the "if-part," but not necessarily vice versa.
  1.22 Acceptance of the "if-part" requires acceptance of the "then-part," but not necessarily vice versa.
  1.23 Instances of an "if-then" statement are implied by the "if-then" statement.
1.3 The rules for categorical reasoning. These rules may be summarized by the following: "Whatever is a member of a general class is also a member of whatever that general class is included in, and is not a member of whatever the general class is excluded from."

A number of cases of reasoning are parallel to strict deduction, but are different in that the generalizations in use as premises do not hold universally under any conceivable circumstance; they have exceptions and limits, not all of which can be specified. To use Waisman's term (1952) they are "open-textured."[3] Reasoning from principles and hypotheses to the world of things, men, and events is inevitably of this sort. Sometimes the exceptions and limits are so far removed that we do not have to worry about them, and in such cases we can proceed as in deduction without fear of going wrong. Sometimes the limits and exceptions are close by, in which case, still approximating the deductive model, we use words like "probable," "likely," "barring unforseen circumstances," etc., in the conclusion.

For an example of the latter case, consider the application of that standard law of economics, "If the supply is constant and the demand for a product decreases, the price will decrease." Two of the limits of the application of this law are within the knowledge of all of us. It is intended to apply to an economy free of government control and to a sector of it that is free of monopolistic control. Mention of these limits will suffice for present purposes, although there are others.

Now let us apply this law to a situation in which there is a decrease in demand for microscopes. Applying the law deductively, we are unalterably committed to a prediction of a price decrease. But it is not wise to be unalterably committed to such a prediction. For one thing, the well-known limits of the law might be breached: the government might decide to maintain the price of microscopes and pay for the destruction of the extras; or a monopoly might be formed which would maintain the price.

But secondly, other things that are not yet explicitly built into the limits might go wrong. The makers of microscopes might form a trade association and decide incorrectly that with good advertising they *can* create a demand much greater than ever before, so that they can afford to raise prices. They therefore raise prices in anticipation of a non-existent demand.

It is because of considerations like this that qualifiers like "probable" must be included in the application of many principles. The application of that law in that situation might be, "It is probable that there will be a lowering of price." But the application would not be this at all if it can be seen that a known limit is breached or that there is some other extenuating circumstance. The point is that the application of such principles should often not be stated any more strongly than this, even though the steps in reasoning parallel those of deduction.

Thus there are two kinds of following: strict necessity and loose following. The critical thinker can do both.

### 2. Judging whether something is an assumption

This topic is complicated because there are various logically-different abilities that go under this title. These can be best approached through an examination of various uses of the word "assumption": the deprecatory use, the concluding use, the premise use, and the presupposition use.

---

[3] Waisman used the term "open texture" to apply to concepts. By applying it to principles I am extending his usage.

2.1 The deprecatory use and the concluding use.

The deprecatory use implies the charge that there is little or no evidence to support a given belief and that the belief is questionable. Here is an example: "You're just assuming that Frank didn't read the assignment." This deprecatory use is often found to be incorporated in the other uses, but sometimes it stands alone. As such its appearance is tantamount to a judgment that the view should perhaps be rejected, or at least be held in abeyance because of lack of support. No further discussion of the evaluation of this kind of assumption-claim is necessary here, since this is a general charge and is covered under discussion of the various other abilities.

In the concluding use, the term "assumption" is used to mark a conclusion, but the deprecatory use is involved too, since the conclusion is implied not to be fully established. Here is an example: "My assumption is that Hissarlik is at the site of Troy" (a statement made at the completion of a presentation of the evidence bearing on the location of Troy). We need not be concerned with discussing whether something is an assumption in the concluding sense; the important question is whether such an assumption is justified and that question is covered elsewhere in this paper.

The first two uses of "assumption" were specified in order to keep them out of the discussion of the next two; the following discussion does not apply to them.

2.2 The premise use.

This kind of assumption stands anterior to a conclusion in a line of reasoning, whether the conclusion be inductive or deductive. To call something an assumption is to say that the conclusion depends upon it, and that if the conclusion is to be accepted, the alleged assumption should also be accepted. Thus the location of assumptions (in this sense) is a useful step in the evaluation of conclusions.

Here are criteria for premise-type assumptions:

2.21 Of the various possible gap-fillers, the alleged assumption should come closest to making the completed argument, proof, or explanation, a satisfactory one. (This criterion is necessary and sufficient.)
  2.211 The simplest gap-filler is ordinarily the one to choose.
  2.212 If there is a more plausible gap-filler among the more complex ones, it should be chosen. Plausibility, however, requires fitting in with existing knowledge—not being a special case.
2.22 Other conditions remaining the same, the state of affairs that is predicted could not occur (or probably would not occur) if the alleged assumption

were false. (This criterion applies only to alleged empirically-necessary assumptions, but for them it is necessary and sufficient.)

2.23 The community of experts in the field would not accept the position, conclusion, or argument, without first believing the assumption to be true. (This criterion is neither necessary nor sufficient, but is a good ground.)

What is a gap-filler? Consider this piece of reasoning:

Since the demand for microscopes has decreased, the price may be expected to decrease.

A gap-filler here would be this principle:

1. When the demand for a commodity decreases, the price will decrease.

It fills a gap in reasoning from the decrease in demand for microscopes to a decrease in price.

It is not the only way to fill the gap, however. Consider these alternatives:

2. When the demand for goods and services decreases, the price decreases.
3. When the demand for optical instruments decreases, the price decreases.
4. When the demand for optical instruments (other than field glasses) decreases, the price decreases.

Since all four of these will fill the gap, it should be clear that being a gap-filler is not by itself a sufficient condition for being an assumption. The simplest gap-filler is ordinarily the one to attribute, thus ruling out *4*. Simplicity might also be a ground for not accepting *2* as the assumption, since there is a conjunction of two things (goods and services) mentioned. But if the prevailing knowledge in economics admits no basis for distinguishing goods from services in the context of this principle, then simplicity is counterbalanced by the need to fit into existing knowledge. Gap-fillers *1* and *2* would then be equally defensible (or indefensible) and either could be called the assumption.

Gap-filler *3* introduces a new twist, talking only about optical instruments. It is as simple as *1* but is not as general. Other things being equal, generality is to be preferred. A system of knowledge is better if it covers more cases. But if the more general gap-filler (*1*) should be false, and the less general one (*3*) true (or more likely to be true), the less general gap-filler is the one to choose. Assumption-finding then is the locating of a gap-filler, the simpler the better, provided that it fits into and contributes to a system of knowledge. The assumption-finder should try to be generous to the person whose assumptions he is locating, generous in that he should try to find the best candidate for participation in a knowledge system. He should not accept a false gap-filler as the assumption until he has searched for one that fits into an acceptable body of knowledge. Put more simply in a way that covers most cases, he should search for one that is true.

While discussing gap-filling it would be well to note that there is one sometimes-used criterion that is inapplicable: *logical* necessity. As exemplified by the four gap-fillers previously discussed, there is no single significant premise-type gap-filler which is logically necessary. It is always *logically* possible (though it may be extremely implausible) to complete an argument in more than one way.[4]

Empirical necessity (2.22) is different. To the extent that empirical statements can be necessary, there can be empirically necessary assumptions. For example, a statement which predicts that the pressure in a fixed cylinder of confined air will increase is assuming that there will be a temperature increase. Since an increase in temperature in that situation *is* necessary for there to be an increase in pressure, the assumption is empirically necessary and can be pinned on the argument with confidence.

Criterion 2.23 mentions the experts. Although their considered opinions can be wrong, they do ordinarily know what fits into their body of knowledge. And they do know what is used successfully in the field to back up arguments and conclusions. So they can ordinarily be expected to know what an argument would need in order to be a good one.

2.3 The presupposition use.

Presuppositions are sentences which must be true for a given statement even to make sense.[5] The claim, "The governor's mistakes have caused our present plight," presupposes that the governor has made mistakes. His not having done so would make nonsense out of either the affirmation or denial of the claim. If the governor has made no mistakes, it does not even make sense to say that his mistakes have caused our plight; nor does it make sense to say that his mistakes have not caused our plight.

Presupposition-finding is useful in avoiding being swayed by false presuppositions (if the governor has made no mistakes, we should be able to react to the presupposition that he has). And presupposition-finding is useful in grasping a verbal picture, and a part or the whole of a theory.

Judging whether something is presupposed by something else is simply a matter of stating the meaning of the "something else."

## 3. Judging whether an observation statement is reliable

An observation statement is a specific description. Over the years, those fields most concerned with accuracy of observation have built up a set of rules for judging the reliability of observation statements. Here is a combined list of principles from the fields of law, history, and science:

---

[4] This point I have developed elsewhere (1961).

[5] This is the sense of the term "presupposition" that was presented by P. F. Strawson (1952).

3.1 Observation statements tend to be more reliable if the *observer:*
  3.11 Was unemotional, alert, and disinterested.
  3.12 Was skilled at observing the sort of thing observed.
  3.13 Had sensory equipment that was in good condition.
  3.14 Has a reputation for veracity.
  3.15 Used precise techniques.
  3.16 Had no preconception about the way the observation would turn out.
3.2 Observation statements tend to be more reliable if the *observation conditions:*
  3.21 Were such that the observer had good access.
  3.22 Provided a satisfactory medium of observation.
3.3 Observation statements tend to be more reliable to the extent that the *statement:*
  3.31 Is close to being a statement of direct observation.
  3.32 Is corroborated.
  3.33 Is corroboratable.
  3.34 Comes from a disinterested source with a reputation for veracity.
3.4 Observation statements, if based on a record, tend to be more reliable if the *record:*
  3.41 Was made at the time of observation.
  3.42 Was made by the person making the statement.
  3.43 Is believed by the person making the statement to be correct—either because he so believed at the time the record was made, or because he believes it was the record-maker's habit to make correct records.
3.5 Observation statements tend to be more reliable than inferences made from them.

### 4. Judging whether a simple generalization is warranted

A simple generalization is a statement which covers a number of instances and holds that they share some trait. For example, that red-headed people tend to have hot tempers is a generalization. It holds that red-headed people share the trait of tending to have hot tempers. A generalization is warranted:

4.1 To the extent that there is a bulk of reliable instances of it. The greater the variability of the population, the greater the bulk needed.
4.2 To the extent that it fits into the larger structure of knowledge.
4.3 To the extent that the selecting of instances is unbiased.
  4.31 A pure random sample is unbiased.
  4.32 A systematic sample is unbiased if a careful investigation suggests that there is not a relevant cycle or trend followed by the sampling procedure.
  4.33 Stratification of a population on relevant variables and unbiased sampling within the strata, is likely to be more efficient than 4.31 or 4.32 alone.
  4.34 An unbiased sampling of clusters of the population and unbiased sampling (or complete enumeration) within the clusters is likely to be an

efficient way of sampling when access to separate individual units is difficult.

4.4 To the extent that there are no counter-instances.

The generalization that red-headed people tend to have hot tempers would be warranted to the extent that there is a large number of reliable instances of red-heads with hot tempers, to the extent that we are able to account for red-heads being hot-tempered, to the extent that our instances of red-heads are picked without bias, and to the extent that there is a lack of reliable instances of red-heads with even tempers.

### 5. Judging whether a hypothesis is warranted

Though the word "hypothesis" is often used to refer to a simple generalization, for purposes of marking off an important kind of statement which has a different relationship to its grounds than a simple generalization I will restrict the word "hypothesis" to the job of referring to this other kind of statement. Under this usage a hypothesis is a statement which is fairly directly related to its support by virtue of its power to *explain* this support, rather than being related by virtue of the support's being composed of instances of the hypothesis, as is the case for simple generalizations.

The hypothesis can be either specific (as in the case in law and usually in history) or it can be general (as is ordinarily the case in physical sciences and the social sciences of economics, sociology, and psychology). Here is an example of a specific hypothesis:

Hissarlik is located at the site of Troy.

And here is an example of a general hypothesis:

The pressure in a liquid varies directly as the depth, assuming the pressure at the surface to be zero.

A hypothesis is warranted to the extent that:

5.1 It explains a bulk and variety of reliable data. If a datum is explained, it can be deduced or loosely derived (in the fashion of the application of principles) from the hypothesis together with established facts or generalizations.

5.2 It is itself explained by a satisfactory system of knowledge.

5.3 It is not inconsistent with any evidence.

5.4 Its competitors are inconsistent with the evidence. This principle is the basis of controlled experiments.

5.5 It is testable. It must be, or have been, possible to make predictions from it.

For purposes of illustration let us consider the bearing of each of the criteria on each hypothesis of the above two examples of hypotheses:

*Explaining a bulk and variety of reliable data* (5.1). Since Hissarlik is only an hour's walk from the sea, the Hissarlik hypothesis explains the reported (in the *Iliad*) ability of the Greeks to go back and forth from Troy several times a day. It explains why there are ruins at Hissarlik. These explained reports, it should be noted, can be derived from the Hissarlik hypotheses together with established facts or generalizations:

H-1. Hissarlik is at the site of Troy.
Hissarlik is one hour's walk from the sea.
People are able to walk back and forth several times a day between places that are one hour's walk apart.
*Therefore,* it is probable that the Greeks were able to go back and forth from Troy to the sea several times daily (the explained fact).

H-2. Hissarlik is at the site of Troy.
A large city when abandoned tends to leave ruins.
*Therefore,* it is probable that there are ruins at Hissarlik (the explained fact).

Of course explaining only those two pieces of evidence is not enough to establish the hypothesis. More evidence of different types must be provided.

The pressure hypothesis explains why water spurts farther from a hole near the bottom of a tank than from a hole in the middle of a tank. It also explains the proportional relationships between the following sets of readings of pressure gauges attached to the supply tank in a water system:

| Distance from top of tank (in ft.) | Pressure reading (in lbs./sq. in.) |
|---|---|
| 0 | 0 |
| 5 | 2.1 |
| 10 | 4.2 |

These data can be derived from the hypothesis together with established facts or generalizations:

P-1. The pressure varies directly as the depth.
The greater the pressure at a hole, the farther the liquid will spurt.
The bottom hole is at a greater depth than the middle hole.
*Therefore,* the water spurts farther from the hole near the bottom (the explained fact).

P-2. The pressure varies directly as the depth.
The depth at 10 ft. is twice that at 5 ft.
*Therefore,* the pressure at 10 ft. (4.2 lbs./sq. in.) is twice that at 5 ft. (2.1 lbs./sq. in.), (the explained fact).

Again the explanation of these data alone does not establish the hypothesis. More explained data of various types are needed.

*Being explained by a satisfactory system of knowledge* (5.2). If the Hissarlik hypothesis could itself be tentatively explained by established facts and generalizations, it would then be more acceptable. For example, suppose it were possible to show that the traits of the Trojans and the facts about the geography, climate, and nearby civilization at the time make it probable that the Trojan city would have developed at Hissarlik at the time that Troy was supposed to have existed. If it were possible to show that, the Hissarlik hypothesis would thereby receive support.

Similarly the pressure hypothesis is supported by showing that it can be explained, and thus derived, as follows:

Pressure in a liquid is the numerical equivalent of the weight of a regular column of liquid extending to the top of the container.
The weight of a column of liquid varies directly with its height.
*Therefore,* the pressure in a liquid varies directly with the depth.

*Not being inconsistent with any evidence* (5·3). The Hissarlik hypothesis would be weakened if no springs could be found in the area of Hissarlik, since the *Iliad* mentions two springs in the area, one hot and one cold. The reasoning might go as follows:

Hissarlik is at the site of Troy.
There were probably at least two springs at Troy, one hot and one cold.
Springs tend to remain in existence over the years.
*Therefore,* it is probable that there are at least two springs at Hissarlik, one hot and one cold.

Note that in using the absence of springs as evidence against the hypothesis, we are assuming that springs tend to remain and that the report of the *Iliad* is reliable. Either of these could be wrong. The less dependable these auxiliary assumptions are, the less dependable is our counter-evidence.

The pressure hypothesis would be weakened by the discovery that water spurted out the same amounts at the middle and the bottom, since the hypothesis implies otherwise. That is, it would be weakened if we did not previously have so much by way of other evidence built up in favor of the hypothesis—so much that in this case, one would have a right to suspect such data.

*Its competitors' being inconsistent with the data* (5·4). A competitor of the Hissarlik hypothesis is the hypothesis that Bunarbashi is at the site of Troy. This competing hypothesis is not consistent with the data that Bunarbashi is a three-hours' walk from the sea and that the Greeks were able to go back and forth several times daily, if we assume that the Greeks walked.

A competitor of the earlier-stated pressure hypothesis might be one to the effect that the pressure increases directly as one gets closer to the surface of the earth. This hypothesis is inconsistent with pressure gauge readings on two independent tanks, one over the other, when the top tank has the pressure gauge at its bottom, and the bottom tank has its gauge at the top. The alternative hypothesis implies that the gauge in the upper tank would give the smaller reading. The data are just the opposite.

A controlled experiment is designed to rule out competing hypotheses by producing data inconsistent with them. When we test the hypothesis that a new fertilizer will increase the growth of corn, we put the fertilizer in a corn patch, develop a companion corn patch, the control, identical in every respect possible except for fertilizer, and watch the results. If there is a difference, the fertilizer hypothesis can explain it. But it would not be explained by heavy rainfall, warm weather, sunlight, etc., since both patches supposedly received the same amount. These alternative hypotheses would justify a prediction of no difference and would thus be inconsistent with the data.

It is, of course, impossible to develop a perfectly controlled experiment, since the perfect isolation and variation of a single variable is not possible. The important thing might be a *combination* of weather and fertilizer, or the important thing might have slipped by unnoticed. But we can still see in the controlled experiment an attempt to approximate the logical goal of eliminating hypotheses by turning up data that is inconsistent with them. The controlled experiment is an efficient way of eliminating hypotheses by this method.

It should be noted that there is an implicit assumption of standard or familiar conditions in the reasoning that leads to judgments of explanation and inconsistency. For example the inconsistency resulting from Bunarbashi's being a three-hour walk from the sea is an inconsistency only if Bunarbashi was a three-hour walk from the sea at the time of the Trojan War. In declaring the data to be inconsistent with the competing hypothesis one is gambling that the sea was not at a significantly different level at that time.

This feature of reasoning from hypotheses fits in with the notion of loose reasoning presented under Aspect #1. There are always possible qualifications when we apply principles to the world of things, men, and events.

*Being testable* (5.5). This is a logical criterion, not a criterion of practicality or even physical possibility. The criterion requires only that it must be possible to *conceive* of what would count as evidence for, and what would count as evidence against, the hypothesis. We have already seen that this is possible for each of our hypotheses. The fact that some conceivable tests are not practically possible is not important so far as this criterion is concerned. A conceivable, though presumably physically impossible, test of the pressure hypothesis would involve swimming to the bottom of the ocean with pressure and depth gauges, recording readings at various points along the way.

An hypothesis that appears untestable is this one: "Airplane crashes are caused by gremlins," since it does not appear to be possible to conceive of something that would count as evidence for and to conceive of something that would count as evidence against the hypothesis. The word "appear" is used deliberately, since the conceiving of evidence for and evidence against would immediately make the hypothesis testable—and would reveal to some extent the meaning of the hypothesis.

Most hypotheses that we consider are testable in this logical sense, so this criterion does not often discredit an hypothesis. But its fulfillment is absolutely essential for hypotheses about the world of things, men, and events.

### 6. Judging whether a theory is warranted

The difference between a theoretic system and an hypothesis of the type we have been considering is that the former is an involved network of relations between concepts, many of which are abstract and technical, while the latter is a simple relation between two or a small number of concepts, often less abstract and technical. Examples of theoretic systems are the kinetic theory of the matter, the atomic theory, Gestalt psychology, the theory of evolution, Keynesian economics, Turner's frontier theory, and classical English grammar. Obviously evaluation of theories is a demanding task. It demands more than we can ordinarily expect of elementary and secondary students. Undergraduates are sometimes better equipped, and graduate students are expected to become equipped to perform this task.

Evaluating theories is comparable to evaluating hypotheses, but much more complex. In general the same criteria apply but on a broader scale. Two modifications should be noted: the addition of the criterion of simplicity and the weakening of the effect of contrary evidence.

The criterion of simplicity calls for choosing the simpler of two competing systems, other things being equal. The classic example of the application of this criterion is the preference of the Copernican system, which considered the sun the center of the universe, to the Ptolemaic system, which looked upon the earth as the center. The Copernican system was simpler since it needed fewer cycles and epicycles to explain the movements of the planets.

Since theories have so many parts, contrary evidence does not usually result in outright rejection, but rather in adjustment to fit the contrary evidence —until the whole system becomes more complex than a competitor. The criterion of simplicity then functions.

Following are the criteria for theoretic systems. There will be brief comments, but no attempt to exemplify the operation of each will be made, because to do so would be a monumental task and would make rather laborious reading for those not versed in the fields chosen. You are invited to provide examples from a theoretic system in a field you know.

A theoretic system is warranted to the extent that:

6.31 It explains a bulk and variety of reliable data. Within the system, furthermore, the less abstract statements should be explained by the more abstract ones.

6.32 It is explained by broader theories. Some theories are so broad already that, with our present state of knowledge, to demand fulfillment of this criterion is often to demand speculation.

6.33 It is not inconsistent with any evidence. As indicated earlier, occasional inconsistency can be handled by adjusting the theory. Sometimes the inconsistency must just be accepted for lack of a better theory, and we say, "The theory does not hold for this kind of case."

6.34 Its competitors are inconsistent with the data. Again a single inconsistency does not destroy a competitor, for it too can be adjusted, but a large number of inconsistencies damage it.

6.35 It is testable. When adjusting a theory to fit the data, people are sometimes tempted to make the theory impregnable by making it untestable. Freudian psychology is sometimes accused of being untestable.

6.36 It is simpler than its rivals. As theories are adjusted to fit new data, they may become extremely complicated, as had happened to the entire Ptolemaic system at the time of Copernicus.

### 7. Judging whether an argument depends on an ambiguity

The ambiguity can appear anywhere in an argument, but most frequently it appears in a shift of meaning from the sense in which the conclusion is proved to a sense in which it is applied. There is such a shift in the following line of reasoning:

> There are people who sincerely believe on religious grounds that medication is wrong. They believe this because they believe that any treatment of human beings with medicine is a violation of their religious principles. "Medication" means "anything intended for the prevention, cure, or alleviation of disease." Since the chlorination of water is intended for the prevention of disease, it is medication. To chlorinate water is thus to violate their religious principles.

The statement, "Chlorination is medication," is proven when the statement has one meaning: "Chlorination is something intended for the prevention, cure, or alleviation of disease." And it is applied with a different meaning: "Chlorination is treatment of human beings with medicine."

Obviously arguments that depend on ambiguities are to be rejected. No criteria can be given that will serve as guides to students in detecting ambiguities, although you can exhort them to be alert with such statements as, "Make sure that the key words are used in the same sense throughout," or "Check the argument using the key word in its ordinary sense, and if it fails, check it using the word in any technical senses that might be employed."

### 8. Judging whether a statement is overvague or overspecific

For the purposes of a given situation, a particular statement might be too vague to provide guidance. In such situations the statement should be rejected or inquired into, since in its condition its truth or falsity is irrelevant.

The statement, "Education has disappeared from the schools" (or, "There is more education in the schools than ever before") is useless in decision-making about curriculum and school finance until the terms, "education," "disappeared," and, "the schools," are clarified. The statements are not specific enough to be tested and applied. They are too vague.

On the other hand, in a war-ravaged country it might be quite meaningful to say that education has disappeared from the schools (since they are now used for hospitals or housing). In this situation, "education," even loosely defined, *has* disappeared from the schools.

This aspect requires consideration of the purpose of the discourse and requires the judgment, "This is (or is not) specific enough for our purpose." If the purpose is to come up with curriculum and budgetary recommendations for a school system long in existence, the statement is not specific enough. If the purpose is to make a report to the leader of a war-ravaged country, it is specific enough.

It might be thought that this aspect of critical thinking is one in which people do not make mistakes. In concrete situations this tends to be true, but in abstract situations it is easy to go wrong by forgetting to put questions and answers in the context of situations with purposes.[6]

### 9. Judging whether an alleged authority is reliable

In order to assess the statements made by an alleged authority, one must appraise his credentials. Certainly other aspects of critical thinking should also be applied, if one is able to do so. But there are times when one must make a judgment about a statement solely on the basis of the credentials of the person making the statement. An alleged authority should be accepted to the extent that:

9.1 He has a good reputation.
9.2 The statement is in his field.
9.3 He was disinterested—that is, he did not knowingly stand to profit by the results of his statements (except that he may have stood to have his reputation affected).
9.4 His reputation could be affected by his statement and he was aware of this fact when he made his statement.
9.5 He studied the matter.
9.6 He followed the accepted procedures in coming to his conclusion (although there are legitimate exceptions to this requirement).
9.7 He was in full possession of his faculties when he made the statement.

---

[6] Crawshay-Williams (1957) develops this point well.

## The Reduction of Problem Identification to the Other Aspects

Different kinds of judgments go under the label, "problem identification":

A. Judging that a want has been identified, as when someone says, "My problem is to learn to appreciate poetry." In this sense the judgment that the speaker has identified his problems is tantamount to the judgment that this is something the speaker, who might also be the judge, really wants to do. Problem identification here is identification of wants, either one's own, or someone else's. If they are one's own introspected wants, then critical thinking is not involved. For a person to know his wants (felt needs) is something that he cannot fail to do.

If they are someone else's wants, then identifying problems is the same as establishing explanatory hypotheses, as is the case for all subconscious wants, one's own or someone else's: for example, "Mark's problem is to get attention." Judging the identification of someone else's problem and of subconscious wants are then critical thinking of a type already discussed—judging hypotheses.

B. Judging that a valuable goal has been selected. Here is such a problem identification: "Our problem in Culver City is to increase respect for law and order." Insofar as that is a statement of an end rather than a means, the judgment that it is an adequate identification of a problem is a value judgment. For reasons indicated earlier, this type of judging, though important, is excluded from this analysis of critical thinking.

C. Judging that a means decision is adequate. For example, if the broader objective were respect for law and order, the following might be a statement of a means decision: "Our problem in Culver City is to establish a youth bureau." The judgment here that the problem has been identified does at least these two things: (1) implies endorsement of the goal of respect for law and order (this part of the judgment then is a value judgment); (2) says that the means selected will facilitate achievement of the goal and that they will be at least more likely to facilitate it than any other course of action, within the limits of existing resources and goals. These limiting goals, by the way, are another place where values are impressed on problem identification. To judge that the problem has been identified is to judge that no unjustified goal violation would take place if the problem were solved.

To apply the means interpretation to our example: it is there implied that establishment of a youth bureau would increase the likelihood of winning respect for law and order, and would be more likely to do so than any other course, such as giving the job to the schools (given the limits of the school budget, the state funds available for youth bureaus, trained personnel available, the goals of the school, etc.).

Judging a means decision is judging the application of a principle and judging the acceptability of the principle. To judge whether a youth bureau in Culver City would result in increased respect for law and order is to judge whether a principle about the effectiveness of youth bureaus, applied to this situation, gives us this statement with sufficient probability; and to judge whether the principle is acceptable. Judging principles comes under judging generalizations, hypotheses, or theories, depending on the principle in question.

In summary, problem identification is many different things and often a combination of them. Elements capable of being treated under the proposed notion of critical thinking are (1) judging the alleged identification of the wants of others and of subconscious wants (explanatory hypotheses), and (2) judging the assertion of a means of reaching a goal (judging the application of principles and judging the principles themselves). Each of these types of judging is treated elsewhere.

## Summary

In this paper you have seen a root notion of critical thinking: the correct assessing of statements; and the presentation and clarification of a list of nine major aspects of critical thinking, which are based upon the root notion. These aspects get at the most important ways people can go wrong in assessing statements and can serve as a statement of elementary and intermediate goals in the teaching of critical thinking.

It has not been the purpose of this paper to suggest how to teach critical thinking, since that would vary so much from one level to another and one subject to another. Perhaps the examples used will suggest teaching ideas to you.

## SELECTED REFERENCES

1. Beck, L. W. Constructions and inferred entities. In H. Feigl & May Brodbeck (Eds.), *Readings in the philosophy of science*. New York: Appleton-Century-Crofts, 1953. Pp. 368–381.
2. Black, M. *Critical thinking*. Englewood Cliffs, N.J.: Prentice-Hall, 1952.
3. Black, M. Definition, presupposition, and assertion. In M. Black, *Problems of analysis*. Ithaca, N.Y.: Cornell Univer. Press, 1954. Pp. 24–45.
4. Black, M. The definition of scientific method. In M. Black, *Problems of analysis*. Ithaca, N.Y.: Cornell Univer. Press, 1954. Pp. 3–23.
5. Black, M. Reasoning with loose concepts, *Dialogue*, 1963, 2 (1), 1–12.
6. Braithwaite, R. B. *Scientific explanation*. London: Cambridge Univer. Press, 1953.

7. Bruner, J. S., Goodnow, Jacqueline J., & Austin, G. A. *A study of thinking.* New York: John Wiley, 1956.
8. Burt, C. The development of reasoning in school children, *J. of exper. Pedagogy,* (1919), 5, 68–77, 121–127.
9. Carnap, R. The interpretation of physics. In H. Feigl & May Brodbeck (Eds.), *Readings in the philosophy of science.* New York: Appleton-Century-Crofts, 1953. Pp. 309–318.
10. Cohen, M. R., & Nagle, E. *An introduction to logic and scientific method.* New York: Harcourt Brace, 1934.
11. Conant, J. B. *On understanding science.* New York: Mentor Books, 1951.
12. Conant, J. B. (Ed.) *Harvard case histories in experimental science.* Cambridge, Mass.: Harvard Univer. Press, 1950–55.
13. Crawshay-Williams, R. *Methods and criteria of reasoning.* London: Routledge and Kegan Paul, 1957.
14. Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests, *Psychol. Bull.,* 1955, 7 (4), 281–302.
15. Dewey, J. *How we think.* Boston: D. C. Heath, 1933.
16. Dray, W. *Laws and explanation in history.* London: Oxford Univer. Press, 1957.
17. Dressel, P. L., & Mayhew, L. B. *General education explorations in evaluation.* Washington, D.C.: American Council on Education, 1954.
18. Ennis, R. H. Assumption-finding. In B. O. Smith & R. H. Ennis (Eds.), *Language and concepts in education.* Chicago: Rand McNally, 1961. Pp. 161–178.
19. Gardiner, P. *The nature of historical explanation.* London: Oxford Univer. Press, 1952.
20. Gasking, D. Causation and recipes, *Mind,* 1956, 64 (256), pp. 479–487.
21. Gottschalk, L. *Understanding history.* New York: Knopf, 1951.
22. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1950.
23. Hall, R. Assuming: One set of positing words, *Phil. Rev.,* 1958, 68 (1), 52–75.
24. Hart, H. L. A., & Honore, A. M. *Causation in the law.* Oxford: Oxford Univer. Press, 1959.
25. Hart, H. L. A. *Definition and theory in jurisprudence.* London: Oxford Univer. Press, 1953.
26. Hempel, C. G. *Fundamentals of concept formation in empirical science.* Chicago: Univer. of Chicago Press, 1952.
27. Hempel, C. G. The function of general laws in history. In H. Feigl & W. Sellars (Eds.), *Readings in philosophical analysis.* New York: Appleton-Century-Crofts, 1949. Pp. 459–471. Also in P. Gardiner (Ed.), *Theories of history.* Glencoe, Ill.: The Free Press, 1959. Pp. 344–356.
28. Henderson, K. B., & Smith, B. O. *A study in the teaching of critical thinking.* Urbana, Ill.: Univer. of Illinois Press (publication date indefinite).
29. Hood, F. C. *Developing the ability to assess the results of thinking.* Urbana, Ill.: University of Illinois Bulletin, 1949.

30. Hospers, J. What is explanation? In A. Flew (Ed.), *Essays in conceptual analysis*. London: Macmillan, 1956. Pp. 94–119.
31. Keynes, J. N. *Formal logic.* (4th ed.) London: Macmillan, 1906.
32. McCarthy, P. J. Sampling. Bulletin No. 15. Ithaca, N.Y.: New York State School of Industrial and Labor Relations, 1951.
33. Morris, C. W. *Foundations of the theory of signs*. Chicago: Univer. of Chicago Press, 1938.
34. Moore, T. V. The reasoning ability of children in the first years of school life, *Studies in psychology and psychiatry from the Catholic University of America*, 1929, 2 (2).
35. Nagel, E. Some issues in the logic of historical analysis. In P. Gardiner (Ed.), *Theories of history*. Glencoe, Ill.: The Free Press, 1959. Pp. 373–385.
36. Robinson, R. *Definition*. London: Oxford Univer. Press, 1954.
37. Scheffler, I. Explanation, prediction, and abstract entities, *British J. for the Phil. of Sci.*, 1956–1957; 7, 293–309.
38. Scriven, M. Definitions, explanations, and theories. In H. Feigl & M. Scriven (Eds.), *Concepts, theories, and the mind-body problem*. Minneapolis: Univer. of Minnesota Press, 1958. Pp. 99–195.
39. Scriven, M. Truisms as the grounds for historical explanation. In P. Gardiner (Ed.), *Theories of history*. Glencoe, Ill.: The Free Press, 1959. Pp. 443–475.
40. Smith, B. O. The improvement of critical thinking, *Progr. Educ.* 1953, 30 (5), 129–134.
41. Smith, B. O. Logic, thinking, and teaching. *Educ. Theor.*, 1957, 7 (4), 225–233.
42. Stevenson, C. L. *Ethics and language*. New Haven: Yale Univer. Press, 1944.
43. Strawson, P. F. *Introduction to logical theory*. London: Methuen, 1952.
44. Toulmin, S. E. *The philosophy of science*. London: Hutchinson, 1953.
45. Toulmin, S. E. Probability. In A. Flew (Ed.), *Essays in conceptual analysis*. London: Macmillan, 1956. Pp. 157–191.
46. Toulmin, S. E. *The uses of argument*. Cambridge: Cambridge Univer. Press, 1958.
47. Urmson, J. O. Parenthetical verbs. In A. Flew (Ed.), *Essays in conceptual analysis*. London: Macmillan, 1956. Pp. 192–212.
48. Valentine, C. W. *Intelligence tests for children*. (6th ed.) London: Methuen, 1958.
49. Waismann, F. Verifiability. In A. Flew (Ed.), *Logic and language*. Oxford: Basil Blackwell & Mott, 1952.
50. Wigmore, J. H. *A students' textbook of the laws of evidence*. Brooklyn, N.Y.: The Foundation Press, 1935.
51. Will, F. H. Generalization and evidence. In M. Black (Ed.), *Philosophical analysis*. Ithaca, N.Y.: Cornell Univer. Press, 1950. Pp. 384–413.

# 24

# The Nature and Nurture of Creative Talent

## — Donald W. MacKinnon

*Why are some men able to make exciting and original contributions within a science or an art while others are bound by existing conventions? This article is based on the Walter Van Dyke Bingham Lecture, given by the author at Yale University in 1962. The substance of the article is derived from research carried on over a period of years at the Institute of Personality Assessment and Research, on the Berkeley campus of the University of California. Three criteria for the identification of creativity, as this is defined in the present study, are given. Also, a description of some of the intellectual and personality characteristics of creative scientists and artists is reported along with findings from an intensive study of the life histories of creative architects. Implications for the nurture of creativity in schools, through the processes of education, are discussed in terms of inferences drawn from the research.*

LET ME say first how deeply appreciative I am of the honor of having been chosen the Walter Van Dyke Bingham Lecturer for 1962. It has for me especial meaning to be provided this opportunity to honor the memory of a man I respected so much and whose work was such a pioneering contribution to that field of psychology to which I have given most of my energies as a psychologist. I am grateful, too, for this opportunity to express to Mrs. Bingham the gratitude of all psychologists for her generosity in establishing

Reprinted from *American Psychologist*, 1962, 17, 484–495, by permission of the author and the American Psychological Association.

this series of annual lectures on the discovery and development of exceptional abilities and capacities. Our literature has been greatly enriched by the lectures which she has made possible.

I should like also to congratulate Yale University for having been chosen this year as the institution to be honored for its contributions to the study of talent, and to thank all those who have made such pleasant arrangements for this occasion.

There is a story, first told I believe by Mark Twain, which, had Dr. Bingham known it, would have been, I am sure, one of his favorites. It is about a man who sought the greatest general who had ever lived. Upon inquiring as to where this individual might be found, he was told that the person he sought had died and gone to Heaven. At the Pearly Gates he informed St. Peter of the purpose of his quest, whereupon St. Peter pointed to a soul nearby. "But that," protested the inquirer, "isn't the greatest of all generals. I knew that person when he lived on earth, and he was only a cobbler." "I know that," replied St. Peter, "but if he had been a general he would have been the greatest of them all."

Dr. Bingham spent his life worrying about cobblers who might have been generals and indeed about all those who fail to become what they are capable of becoming because neither they nor others recognize their potentialities and nourish their realization. Dr. Bingham was one of the first to insist that it is not enough to recognize creative talent after it has come to expression. He reminded us that it is our task as psychologists and as educators either through our insights or through the use of validated predictors to discover talent when it is still potential and to provide that kind of social climate and intellectual environment which will facilitate its development and expression.

Whatever light I shall be able to shed on the nature and nurture of creative talent comes in the main from findings of researches carried on during the last six years in the Institute of Personality Assessment and Research on the Berkeley campus of the University of California, and supported in large part by the Carnegie Corporation of New York.

In undertaking such a study one of our first tasks was to decide what we would consider creativity to be. This was necessary, first, because creativity has been so variously described and defined; and, second, because only when we had come to agreement as to how we would conceive creativity would we be in a position to know what kinds of persons we would want to study.

We came easily to agreement that true creativeness fulfills at least three conditions. It involves a response or an idea that is novel or at the very least statistically infrequent. But novelty or originality of thought or action, while a necessary aspect of creativity, is not sufficient. If a response is to lay claim to being a part of the creative process, it must to some extent be adaptive to, or of, reality. It must serve to solve a problem, fit a situation, or accomplish some recognizable goal. And, thirdly, true creativeness involves a sustaining of the original insight, an evaluation and elaboration of it, a developing of it to the full.

Creativity, from this point of view, is a process extended in time and characterized by originality, adaptiveness, and realization. It may be brief, as in a musical improvisation, or it may involve a considerable span of years, as was required for Darwin's creation of the theory of evolution.

The acceptance of such a conception of creativity had two important consequences for our researches. It meant that we would not seek to study creativity while it was still potential but only after it had been realized and had found expression in clearly identifiable creative products—buildings designed by architects, mathematical proofs developed by mathematicians, and the published writings of poets and novelists. Our conception of creativity forced us further to reject as indicators or criteria of creativeness the performance of individuals on so-called tests of creativity. While tests of this sort, that require that the subject think, for example, of unusual uses for common objects and the consequences of unusual events, may indeed measure the infrequency or originality of a subject's ideas in response to specific test items, they fail to reveal the extent to which the subject faced with real life problems is likely to come up with solutions that are novel and adaptive and which he will be motivated to apply in all of their ramifications.

Having thus determined that we would limit our researches to the study of persons who had already demonstrated a high level of creative work, we were still confronted with the problem of deciding from which fields of creative endeavor we would seek to recruit our subjects.

The fields which we finally sampled were those of creative writing, architecture, mathematics, industrial research, physical science, and engineering.

If one considers these activities in relation to the distinction often made between artistic and scientific creativity, it may be noted that we have sampled both of these domains as well as overlapping domains of creative striving which require that the practitioner be at one and the same time both artist and scientist.

Artistic creativity, represented in our studies by the work of poets, novelists, and essayists, results in products that are clearly expressions of the creator's inner states, his needs, perceptions, motivations, and the like. In this type of creativity, the creator externalizes something of himself into the public field.

In scientific creativity, the creative product is unrelated to the creator as a person, who in his creative work acts largely as a mediator between externally defined needs and goals. In this kind of creativeness, the creator, represented in our studies by industrial researchers, physical scientists, and engineers, simply operates on some aspect of his environment in such a manner as to produce a novel and appropriate product, but he adds little of himself or of his style as a person to the resultant.

Domains of creative striving in which the practitioner must be both artist and scientist were represented in our researches by mathematicians and architects. Mathematicians contribute to science, yet in a very real sense their important creative efforts are as much as anything else personal cosmologies in which they express themselves as does the artist in his creations. So, too, in architecture, creative products are both an expression of the architect and thus a very personal product, and at the same time an impersonal meeting of the demands of an external problem.

If in reporting the findings of our researches I draw most heavily upon data obtained from our study of architects (MacKinnon, 1962), it is for two reasons. First, it is the study for which, in collaboration with Wallace B. Hall, I have assumed primary responsibility. Second, it is in architects, of all our samples, that we can expect to find what is most generally characteristic of creative persons. Architecture, as a field of creative endeavor, requires that the successful practitioner be both artist and scientist—artist in that his designs must fulfill the demands of "Delight," and scientist in that they must meet the demands of "Firmnesse" and "Commodity," to use the words of Sir Henry Wotton (1624). But surely, one can hardly think that the requirements of effective architecture are limited to these three demands. The successful and effective architect must, with the skill of a juggler, combine, reconcile, and exercise the diverse skills of businessman, lawyer, artist, engineer, and advertising man, as well as those of author and journalist, psychiatrist, educator, and psychologist. In what other profession can one expect better to observe the multifarious expressions of creativity?

It should be clear that any attempt to discover the distinguishing traits of creative persons can succeed only in so far as some group of qualified experts can agree upon who are the more and who are the less creative workers in a given field of endeavor. In our study of architects we began by asking a panel of experts—five professors of architecture, each working independently—to nominate the 40 most creative architects in the United States. All told they supplied us with 86 names instead of the 40 they would have mentioned had there been perfect agreement among them. While 13 of the 86 architects were nominated by all five panel members, and 9 nominated by four, 11 by three, and 13 by two, 40 were individual nominations, each proposed by a single panel member.

The agreement among experts is not perfect, yet far greater than one might have expected. Later we asked 11 editors of the major American architectural journals, *Architectural Forum*, *Architectural Record*, the *Journal of* *the American Institute of Architects,* and *Progressive Architecture,* to rate the creativity of the 64 of the nominated architects whom we invited to participate in the study. Still later we asked the 40 nominated creative architects who actually accepted our invitation to be studied to rate the creativity of the invited 64 architects, themselves included. Since the editors' ratings of the creativity of the architects correlated +.88 with the architects' own ratings, it is clear that under certain conditions and for certain groups it is possible to obtain remarkable agreement about the relative creativeness of individual members of a profession and thus meet the first requirement for an effective study of creative persons.

A second requirement for the successful establishment of the traits of creative individuals is their willingness to make themselves available for study. Our hope was to win the cooperation of each person whom we invited to participate in the research, but as I have already indicated in the case of the architects, to obtain 40 acceptances, 64 invitations had to be sent out.

The invitation to this group, as to all the creative groups which we have studied, was to come to Berkeley for a weekend of intensive study in the Institute of Personality Assessment and Research. There, in groups of ten, they have been studied by the variety of means which constitute the assessment method—by problem solving experiments; by tests designed to discover what a person does not know or is unable or unwilling to reveal about himself; by tests and questionnaires that permit a person to manifest various aspects of his personality and to express his attitudes, interest, and values; by searching interviews that cover the life history and reveal the present structure of the person; and by specially contrived social situations of a stressful character which call for the subject's best behavior in a socially defined role.

The response of creative persons to the invitation to reveal themselves under such trying circumstances has varied considerably. At the one extreme there have been those who replied in anger at what they perceived to be the audacity of psychologists in presuming to study so ineffable and mysterious a thing as the creative process and so sensitive a being as a creative person. At the other extreme were those who replied courteously and warmheartedly, welcoming the invitation to be studied, and manifesting even an eagerness to contribute to a better understanding of the creative person and the creative process.

Here we were face to face with a problem that plagues us in all our researches: Are those who are willing to be assessed different in important ways from those who refuse? With respect to psychological traits and characteristics we can never know. But with respect to differences in creativeness, if any, between the 40 who accepted and the 24 who declined our invitation, we know that the two groups are indistinguishable. When the nominating panel's ratings of creativity were converted to standard scores and the means for the 24 versus the 40 were compared, they were found to be identical. When the editors' ratings were similarly converted to standard scores, the mean for the nonassessed group was slightly higher (51.9) than for the assessed sample (48.7), but the difference is not statistically significant.

Certainly we cannot claim to have assessed the 40 most creative architects in the country, or the most creative of any of the groups we have studied; but it is clear that we have studied a highly creative group of architects indistinguishable in their creativity from the group of 24 who declined to be studied, and so with the other groups too.

A third requirement for the successful determination of the traits of highly creative persons in any field of endeavor is that the profession be widely sampled beyond those nominated as most creative, for the distinguishing characteristics of the restricted sample might well have nothing to do with their creativeness. Instead they might be traits characterizing all members of the profession whether creative or not, distinguishing the professional group as a whole but in no sense limited or peculiar to its highly creative members. In the case of the architects, to use them once again as an example, two additional samples were recruited for study, both of which matched the highly creative sample (whom I shall now call Architects I) with respect to age and geographic location of practice. The first supplementary sample (Architects II) had had at least two years of work experience and association with one of the originally nominated creative architects. The second additional sample (Architects III) was composed of architects who had never worked with any of the nominated creatives.

By selecting three samples in this manner, we hoped to tap a range of talent sufficiently wide to be fairly representative of the profession as a whole; and we appear to have succeeded. The mean rating of creativity for each of the three groups—the ratings having been made on a seven point scale by six groups of architects and experts on architecture—was for Architects I, 5.46; for Architects II, 4.25; and for Architects III, 3.54, the differences in mean ratings between each group being statistically highly significant.

So much for method and research design. I turn now to a discussion of the nature of creative talent as it has been revealed to us in our researches.

Persons who are highly creative are inclined to have a good opinion of themselves, as evidenced by the large number of favorable adjectives which they use in self-description and by the relatively high scores they earn on a scale which measures basic acceptance of the self. Indeed, there is here a paradox, for in addition to their favorable self-perceptions the very basic self-acceptance of the more creative persons often permits them to speak more frankly and thus more critically and in unusual ways about themselves. It is clear, too, that the self-images of the more creative differ from the self-images of the less creative. For example, Architects I, in contrast to Architects II and III, more often describe themselves as inventive, determined, independent, individualistic, enthusiastic, and industrious. In striking contrast Architects II and III more often than Architects I describe themselves as responsible, sincere, reliable, dependable, clear thinking, tolerant, and understanding. In short, where creative architects more often stress their inventiveness, independence, and individuality, their enthusiasm, determination, and industry, less creative members of the profession are impressed by their virtue and good character and by their rationality and sympathetic concern for others.

The discrepancies between their descriptions of themselves as they are and as they would ideally be are remarkably alike for all architects regardless of their level of creativeness. All three groups reveal themselves as desiring more personal attractiveness, self-confidence, maturity, and intellectual competence, a higher level of energy, and better social relations. As for differences, however, Architects I would ideally be more sensitive, while both Architects II and III wish for opposites if not incompatibles; they would ideally be more original but at the same time more self-controlled and disciplined.

As for the relation between intelligence and creativity, save for the mathematicians where there is a low positive correlation between intelligence and the level of creativeness, we have found within our creative samples essentially zero relationship between the two variables, and this is not due to a narrow restriction in range of intelligence. Among creative architects who have a mean score of 113 on the Terman Concept Mastery Test (1956), individual scores range widely from 39 to 179, yet scores on this measure of intelligence correlate −.08 with rated creativity. Over the whole range of intelligence and creativity there is, of course, a positive relationship between the two variables. No feeble-minded subjects have shown up in any of our creative groups. It is clear, however, that above a certain required minimum level of intelligence which varies from field to field and in some instances may be surprisingly low, being more intelligent does not guarantee a corresponding increase in creativeness. It just is not true that the more intelligent person is necessarily the more creative one.

In view of the often asserted close association of genius with insanity it is also of some interest to inquire into the psychological health of our creative subjects. To this end we can look at their profiles on the Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway & McKinley, 1945), a test originally developed to measure tendencies toward the major psychiatric disturbances that man is heir to: depression, hysteria, paranoia, schizophrenia, and the like. On the eight scales which measure the strength of these dispositions in the person, our creative subjects earn scores which, on the average, are some 5 to 10 points above the general populations' average score of 50. It must be noted, however, that elevated scores of this degree on these scales do not have the same meaning for the personality functioning of persons who, like our subjects, are getting along well in their personal lives and professional careers, that they have for hospitalized patients. The manner in which creative subjects describe themselves on this test as well as in the life history psychiatric interview is less suggestive of psychopathology than it is of good intellect, complexity and richness of personality, general lack of defensiveness, and candor in self-description—in other words, an openness to experience and especially to experience of one's inner life. It must also be noted, however, that in the self-reports and in the MMPI profiles of many of our creative subjects, one can find rather clear evidence of psychopathology, but also evidence of adequate control mechanisms, as the success with which they live their productive and creative lives testifies.

However, the most striking aspect of the MMPI profiles of all our male creative groups in an extremely high peak on the *Mf* (femininity) scale. This tendency for creative males to score relatively high on femininity is also demonstrated on the Fe (femininity) scale of the California Psychological Inventory (CPI) (Gough, 1957) and on the masculinity-femininity scale of the Strong Vocational Interest Blank (Strong, 1959). Scores on the latter scale (where high score indicates more masculinity) correlate −.49 with rated creativity.

The evidence is clear: The more creative a person is the more he reveals an openness to his own feelings and emotions, a sensitive intellect and understanding self-awareness, and wide-ranging interests including many which in the American culture are thought of as feminine. In the realm of sexual identification and interests, our creative subjects appear to give more expression to the feminine side of their nature than do less creative persons. In the language of the Swiss psychologist, Carl G. Jung (1956), creative persons are not so completely identified with their masculine *persona* roles as to blind themselves to or to deny expression to the more feminine traits of the *anima.* For some, to be sure, the balance between masculine and feminine traits, interests, and identification, is a precarious one, and for several of our subjects it would appear that their presently achieved reconciliation of these opposites of their nature has been barely effected and only after considerable psychic stress and turmoil.

The perceptiveness of the creative and his openness to richness and complexity of experience is strikingly revealed on the Barron-Welsh Art Scale of the Welsh Figure Preference Test (Welsh, 1959), which presents to the subject a set of 62 abstract line drawings which range from simple and symmetrical figures to complex and asymmetrical ones. In the original study (Barron & Welsh, 1952) which standardized this scale, some 80 painters from New York, San Francisco, New Orleans, Chicago, and Minneapolis showed a marked preference for the complex and asymmetrical, or, as they often referred to them, the vital and dynamic figures. A contrasting sample of nonartists revealed a marked preference for the simple and symmetrical drawings.

All creative groups we have studied have shown a clear preference for the complex and asymmetrical, and in general the more creative a person is the stronger is this preference. Similarly, in our several samples, scores on an Institute scale which measures the preference for perceptual complexity are significantly correlated with creativity. In the sample of architects the correlation is +.48.

Presented with a large selection of one-inch squares of varicolored posterboard and asked to construct within a 30-minute period a pleasing, completely filled-in 8″ × 10″ mosaic (Hall, 1958), some subjects select the fewest colors possible (one used only one color, all white) while others seek to make order out of the largest possible number, using all of the 22 available colors. And, again citing results from the architects, there is a significant though low positive correlation of +.38 between the number of colors a subject chooses and his creativity as rated by the experts.

If one considers for a moment the meaning of these preferences on the art scale, on the mosaic test, and on the scale that measures preference for perceptual complexity, it is clear that creative persons are especially disposed to admit complexity and even disorder into their perceptions without being made anxious by the resulting chaos. It is not so much that they like disorder per se, but that they prefer the richness of the disordered to the stark barrenness of the simple. They appear to be challenged by disordered multiplicity which arouses in them a strong need which in them is serviced by a superior capacity to achieve the most difficult and far-reaching ordering of the richness they are willing to experience.

The creative person's openness to experience is further revealed on the Myers-Briggs Type Indicator (Myers, 1958), a test based largely upon Carl G. Jung's (1923) theory of psychological functions and types.

Employing the language of the test, though in doing so I oversimplify both it and the theory upon which it is based, one might say that whenever a person uses his mind for any purpose, he performs either an act of perception (he becomes aware of something) or an act of judgment (he comes to a conclusion about something). And most persons tend to show a rather consistent preference for and greater pleasure in one or the other of these, preferring either to perceive or to judge, though every one both perceives and judges.

An habitual preference for the judging attitude may lead to some prejudging and at the very least to the living of a life that is orderly, controlled, and carefully planned. A preference for the perceptive attitude results in a life that is more open to experience both from within and from without, and characterized by flexibility and spontaneity. A judging type places more emphasis upon the control and regulation of experience, while a perceptive type is inclined to be more open and receptive to all experience.

The majority of our creative writers, mathematicians, and architects are perceptive types. Only among research scientists do we find the majority to be judging types, and even in this group it is interesting to note that there is a positive correlation (+.25) between a scientist's preference for perception and his rated creativity as a scientific researcher. For architects, preference for perception correlates +.41 with rated creativity.

The second preference measured by the Type Indicator is for one of two types of perception: sense perception or sensation, which is a direct becoming aware of things by way of the senses versus intuitive perception or intuition, which is an indirect perception of the deeper meanings and possibilities inherent in things and situations. Again, everyone senses and intuits, but preliminary norms for the test suggest that in the United States three out of four persons show a preference for sense perception, concentrating upon immediate sensory experience and centering their attention upon existing facts. The one out of every four who shows a preference for intuitive perception, on the other hand, looks expectantly for a bridge or link between that which is given and present and that which is not yet thought of, focusing habitually upon possibilities.

One would expect creative persons not to be bound to the stimulus and the object but to be ever alert to the as-yet-not-realized. And that is precisely the way they show themselves to be on the Type Indicator. In contrast to an estimated 25% of the general population who are intuitive, 90% of the creative writers, 92% of the mathematicians, 93% of the research scientists, and 100% of the architects are intuitive as measured by this test.

In judging or evaluating experience, according to the underlying Jungian theory of the test, one makes use of thought or of feeling; thinking being a logical process aimed at an impersonal fact-weighing analysis, while feeling is a process of appreciation and evaluation of things that gives them a personal and subjective value. A preference for thinking or for feeling appears to be less related to one's creativity as such than to the type of materials or concepts with which one deals. Of our creative groups, writers prefer feeling, mathematicians, research scientists, and engineers prefer thinking, while architects split fifty-fifty in their preference for one or the other of the two functions.

The final preference in Jungian typology and on the test is the well-known one between introversion and extraversion. Approximately two-thirds of all our creative groups score as introverts, though there is no evidence that introverts as such are more creative than extraverts.

Turning to preferences among interests and values, one would expect the highly creative to be rather different from less creative people, and there is clear evidence that they are.

On the Strong Vocational Interest Blank, which measures the similarity of a person's expressed interests with the known interests of individuals successful in a number of occupations and professions, all of our creative subjects have shown, with only slight variation from group to group, interests similar to those of the psychologist, author-journalist, lawyer, architect, artist, and musician, and interests unlike those of the purchasing agent, office man, banker, farmer, carpenter, veterinarian and interestingly enough, too, policeman and mortician. Leaving aside any consideration of the specific interests thus revealed we may focus our attention on the inferences that may be drawn from this pattern of scores which suggest that creative persons are relatively uninterested in small details, or in facts for their own sake, and more concerned with their meanings and implications, possessed of considerable cognitive flexibility, verbally skillful, interested in communicating with others and accurate in so doing, intellectually curious, and relatively disinterested in policing either their own impulses and images or those of others.

On the Allport-Vernon-Lindzey Study of Values (1951), a test designed to measure in the individual the relative strength of the six values of men as these values have been conceptualized and described by the German psychologist and educator, Eduard Spranger (1928), namely, the theoretical, economic, esthetic, social, political, and religious values, all of our creative groups have as their highest values the theoretical and the esthetic.

For creative research scientists the theoretical value is the highest, closely followed by the esthetic. For creative architects the highest value is the esthetic, with the theoretical value almost as high. For creative mathematicians, the two values are both high and approximately equally strong.

If, as the authors of the test believe, there is some incompatibility and conflict between the theoretical value with its cognitive and rational concern with truth and the esthetic value with its emotional concern with form and beauty, it would appear that the creative person has the capacity to tolerate the tension that strong opposing values create in him, and in his creative striving he effects some reconciliation of them. For the truly creative person it is not sufficient that problems be solved, there is the further demand that the solutions be elegant. He seeks both truth and beauty.

A summary description of the creative person—especially of the creative architect—as he reveals himself in his profile on the California Psychological Inventory (Gough, 1957) reads as follows:

> He is dominant (Do scale); possessed of those qualities and attributes which underlie and lead to the achievement of social status (Cs); poised, spontaneous, and self-confident in personal and social interaction (Sp); though not of an especially sociable or participative temperament (low Sy); intelligent, outspoken, sharp-witted, demanding, aggressive, and self-centered; persuasive and verbally fluent, self-confident and self-assured (Sa); and relatively uninhibited in expressing his worries and complaints (low Wb).
>
> He is relatively free from conventional restraints and inhibitions (low So and Sc), not preoccupied with the impression which he makes on others and thus perhaps capable of great independence and autonomy (low Gi), and relatively ready to recognize and admit self-views that are unusual and unconventional (low Cm).
>
> He is strongly motivated to achieve in situations in which independence in thought and action are called for (Ai). But, unlike his less creative colleagues, he is less inclined to strive for achievement in settings where conforming behavior is expected or required (Ac). In efficiency and steadiness of intellectual effort (Ie), however, he does not differ from his fellow workers.
>
> Finally, he is definitely more psychologically minded (Py), more flexible (Fx), and possessed of more femininity of interests (Fe) than architects in general.

There is one last finding that I wish to present, one that was foreshadowed by a discovery of Dr. Bingham in one of his attempts to study creativity. The subject of his study was Amy Lowell, a close friend of his and Mrs. Bingham's, with whom he discussed at length the birth and growth of her poems, seeking insight into the creative processes of her mind. He also administered to her a word association test and "found that she gave a higher proportion of unique responses than those of any one outside a mental institution" (Bingham, Millicent Todd, 1953, p. 11). We, too, administered a word association test to our subjects and found the unusualness of mental associations one of the best predictors of creativity, and especially so when associations given by no more than 1% to 10% of the population, using the Minnesota norms (Russell & Jenkins, 1954), are weighted more heavily than those given by less than 1% of the population. Among architects, for example, this weighted score is for Architects I, 204; Architects II, 128; and Architects III, 114; while for the total sample this measure of unusualness of mental associations correlates +.50 with rated creativity.

And Dr. Bingham, like us, found that there are certain hazards in attempting to study a creative poet. His searchings were rewarded by a poem Amy Lowell later wrote which was first entitled "To the Impudent Psychologist" and published posthumously with the title "To a Gentleman who wanted to see the first drafts of my poems in the interest of psychological research into the workings of the creative mind." We, I must confess, were treated somewhat less kindly by one of our poets who, after assessment, published an article entitled "My Head Gets Tooken Apart" (Rexroth, 1959).

Having described the overall design of our studies, and having presented a selection of our findings which reveal at least some aspects of the nature of creative talent, I turn now, but with considerably less confidence, to the question as to how we can early identify and best encourage the development of creative potential. Our findings concerning the characteristics of highly creative persons are by now reasonably well established, but their implications for the nurture of creative talent are far from clear.

It is one thing to discover the distinguishing characteristics of mature, creative, productive individuals. It is quite another matter to conclude that the traits of creative persons observed several years after school and college characterized these same individuals when they were students. Nor can we be certain that finding these same traits in youngsters today will identify those with creative potential. Only empirical, longitudinal research, which we do not yet have, can settle such issues. Considering, however, the nature of the traits which discriminate creative adults from their noncreative peers, I would venture to guess that most students with creative potential have personality structures congruent with, though possibly less sharply delineated than, those of mature creatives.

Our problem is further complicated by the fact that though our creative subjects have told us about their experiences at home, in school, and in college, and about the forces and persons and situations which, as they see it, nurtured their creativeness, these are, after all, self-reports subject to the misperceptions and self-deceptions of all self-reports. Even if we were to assume that their testimony is essentially accurate we would still have no assurance that the conditions in the home, in school, and society, the qualities of interpersonal relations between instructor and student, and the aspects of the teaching-learning process which would appear to have contributed to creative development a generation ago would facilitate rather than inhibit creativity if these same factors were created in today's quite different world and far different educational climate.

In reporting upon events and situations in the life histories of our subjects which appear to have fostered their creative potential and independent spirit, I shall again restrict myself to architects. One finds in their histories a number of circumstances which, in the early years, could well have provided an opportunity as well as the necessity for developing the secure sense of personal autonomy and zestful commitment to their profession which so markedly characterize them.

What appears most often to have characterized the parents of these future creative architects was an extraordinary respect for the child and confidence in his ability to do what was appropriate. Thus they did not hesitate to grant him rather unusual freedom in exploring his universe and in making decisions for himself—and this early as well as late. The expectation of the parent that the child would act independently but reasonably and responsibly appears to have contributed immensely to the latter's sense of personal autonomy which was to develop to such a marked degree.

The obverse side of this was that there was often a lack of intense closeness with one or both of the parents. Most often this appeared in relation to the father rather than to the mother, but often it characterized the relationship with both parents. There were not strong emotional ties of either a positive or a negative sort between parent and child, but neither was there the type of relationship that fosters overdependency nor the type that results in severe rejection. Thus, if there was a certain distance in the relationship between child and parent, it had a liberating effect so far as the child was concerned. If he lacked something of the emotional closeness which some children experience with their parents, he was also spared that type of psychological exploitation that is so frequently seen in the life histories of clinical patients.

Closely related to this factor of some distance between parent and child were ambiguities in identification with the parents. In place of the more usual clear identification with one parent, there was a tendency for the architects to have identified either with both parents or with neither. It was not that the child's early milieu was a deprived one so far as models for identification and the promotion of ego ideals were concerned. It was rather that the larger familial sphere presented the child with a plentiful supply of diverse and effective models—in addition to the mother and father, grandfathers, uncles, and others who occupied prominent and responsible positions within their community—with whom important identifications could be made. Whatever the emotional interaction between father and son, whether distant, harmonious, or turbulent, the father presented a model of effective and resourceful behavior in an exceptionally demanding career. What is perhaps more significant, though, is the high incidence of distinctly autonomous mothers among families of the creative architects, who led active lives with interests and sometimes careers of their own apart from their husbands'.

Still other factors which would appear to have contributed to the development of the marked personal autonomy of our subjects were the types of discipline and religious training which they received, which suggest that within the family there existed clear standards of conduct and ideas as to what was right and wrong but at the same time an expectation if not requirement of active exploration and internalization of a framework of personal conduct. Discipline was almost always consistent and predictable. In most cases there were rules, family standards, and parental injunctions which were known explicitly by the children and seldom infringed. In nearly half the cases, corporal punishment was not employed and in only a few instances was the punishment harsh or cruel.

As for religious practices, the families of the creative architects showed considerable diversity, but what was most widely emphasized was the development of personal ethical codes rather than formal religious practices. For one-third of the families formal religion was important for one parent or for both, but in two-thirds of the families formal religion was either unimportant or practiced only perfunctorily. For the majority of the families, in which emphasis was placed upon the development of one's own ethical code, it is of interest to inquire into the values that were most stressed. They were most often values related to integrity (e.g., forthrightness, honesty, respect for others), quality (e.g., pride, diligence, joy in work, development of talent), intellectual and cultural endeavor, success and ambition, and being respectable and doing the right thing.

The families of the more creative architects tended to move more frequently, whether within a single community, or from community to community, or even from country to country. This, combined with the fact that the more creative architects as youngsters were given very much more freedom to roam and to explore widely, provided for them an enrichment of experience both cultural and personal which their less creative peers did not have.

But the frequent moving appears also to have resulted frequently in some estrangement of the family from its immediate neighborhood. And it is of interest that in almost every case in which the architect reported that his family differed in its behavior and values from those in the neighborhood, the family was different in showing greater cultural, artistic, and intellectual interests and pursuits.

To what extent this sort of cultural dislocation contributed to the frequently reported experiences of aloneness, shyness, isolation, and solitariness during childhood and adolescence, with little or no dating during adolescence, or to what extent these experiences stemmed from a natural introversion of interests and unusual sensitivity, we cannot say. They were doubtless mutually reinforcing factors in stimulating the young architect's awareness of his own inner life and his growing interest in his artistic skills and his ideational, imaginal, and symbolic processes.

Almost without exception, the creative architects manifested very early considerable interest and skill in drawing and painting. And also, with almost no exception, one or both of the parents were of artistic temperament and considerable skill. Often it was the mother who in the architect's early years fostered his artistic potentialities by her example as well as by her instruction. It is especially interesting to note, however, that while the visual and artistic abilities and interests of the child were encouraged and rewarded, these interests and abilities were, by and large, allowed to develop at their own speed, and this pace varied considerably among the architects. There was not an anxious concern on the part of the parents about the skills and abilities of the child. What is perhaps most significant was the widespread definite lack of strong pressures from the parents toward a particular career. And this was true both for pressures away from architecture as well as for pressures toward architecture by parents who were themselves architects.

The several aspects of the life history which I have described were first noted by Kenneth Craik in the protocols for the highly creative Architects I. Subsequently, in reading the protocols for Architects II and III as well as Architects I, a credit of one point for the presence of each of the factors was assigned and the total for each person taken as a score. The correlation of these life history scores with rated creativity of the architects is +.36, significant beyond the .005 level of confidence.

And now I turn finally to a consideration of the implications of the nature of creative talent for the nurturing of it in school and college through the processes of education.

Our findings concerning the relations of intelligence to creativity suggest that we may have overestimated in our educational system the role of intelligence in creative achievement. If our expectation is that a child of a given intelligence will not respond creatively to a task which confronts him, and especially if we make this expectation known to the child, the probability that he will respond creatively is very much reduced. And later on, such a child, now grown older, may find doors closed to him so that he is definitely excluded from certain domains of learning. There is increasing reason to believe that in selecting students for special training of their talent we may have overweighted the role of intelligence either by settting the cutting point for selection on the intellective dimension too high or by assuming that regardless of other factors the student with the higher IQ is the more promising one and should consequently be chosen. Our data suggest, rather, that if a person has the minimum of intelligence required for mastery of a field of knowledge, whether he performs creatively or banally in that field will be crucially determined by nonintellective factors. We would do well then to pay more attention in the future than we have in the past to the nurturing of those nonintellective traits which in our studies have been shown to be intimately associated with creative talent.

There is the openness of the creative person to experience both from within and from without which suggests that whether we be parent or teacher we should use caution in setting limits upon what those whom we are nurturing experience and express.

Discipline and self-control are necessary. They must be learned if one is ever to be truly creative, but it is important that they not be overlearned. Furthermore, there is a time and place for their learning, and having been learned they should be used flexibly, not rigidly or compulsively.

If we consider this specifically with reference to the attitudes of perceiving and judging, everyone must judge as well as perceive. It is not a matter of using one to the exclusion of the other, but a question of how each is used and which is preferred. The danger for one's creative potential is not the judging or evaluating of one's experience but that one prejudges, thus excluding from perception large areas of experience. The danger in all parental instruction, as in all academic instruction, is that new ideas and new possibilities of action are criticized too soon and too often. Training in criticism is obviously important and so widely recognized that I need not plead its case. Rather I would urge that, if we wish to nurture creative potential, an equal emphasis be placed on perceptiveness, discussing with our students as well as with our children, at least upon occasion, the most fantastic of ideas and possibilities. It is the duty of parents to communicate and of professors to profess what they judge to be true, but it is no less their duty by example to encourage in their children and in their students an openness to all ideas and especially to those which most challenge and threaten their own judgments.

The creative person, as we have seen, is not only open to experience, but intuitive about it. We can train students to be accurate in their perceptions, and this, too, is a characteristic of the creative. But can we train them to be intuitive, and if so how?

I would suggest that rote learning, learning of facts for their own sake, repeated drill of material, too much emphasis upon facts unrelated to other facts, and excessive concern with memorizing, can all strengthen and reinforce sense perception. On the other hand, emphasis upon the transfer of training from one subject to another, the searching for common principles in terms of which facts from quite different domains of knowledge can be related, the stressing of analogies, and similies, and metaphors, a seeking for symbolic equivalents of experience in the widest possible number of sensory and imaginal modalities, exercises in imaginative play, training in retreating from the facts in order to see them in larger perspective and in relation to more aspects of the larger context thus achieved—these and still other emphases in learning would, I believe, strengthen the disposition to intuitive perception as well as intuitive thinking.

If the widest possible relationships among facts are to be established, if the structure of knowledge (Bruner, 1960) is to be grasped, it is necessary that the student have a large body of facts which he has learned as well as a large array of reasoning skills which he has mastered. You will see, then, that what I am proposing is not that in teaching one disdain acute and accurate sense perception, but that one use it to build upon, leading the student always to an intuitive understanding of that which he experiences.

The independence of thought and action which our subjects reveal in the assessment setting appears to have long characterized them. It was already manifest in high school, though, according to their reports, tending to increase in college and thereafter.

In college our creative architects earned about a B average. In work and courses which caught their interest they could turn in an A performance, but in courses that failed to strike their imagination, they were quite willing to do no work at all. In general, their attitude in college appears to have been one of profound skepticism. They were unwilling to accept anything on the mere say-so of their instructors. Nothing was to be accepted on faith or because it had behind it the voice of authority. Such matters might be accepted, but only after the student on his own had demonstrated their validity to himself. In a sense, they were rebellious, but they did not run counter to the standards out of sheer rebelliousness. Rather, they were spirited in their disagreement and one gets the impression that they learned most from those who were not easy with them. But clearly many of them were not easy to take. One of the most rebellious, but, as it turned out, one of the most creative, was advised by the Dean of his School to quit because he had no talent; and another, having been failed in his design dissertation which attacked the stylism of the faculty, took his degree in the art department.

These and other data should remind all of us who teach that creative students will not always be to our liking. This will be due not only to their independence in situations in which nonconformity may be seriously disruptive of the work of others, but because, as we have seen, more than most they will be experiencing large quantities of tension produced in them by the richness of their experience and the strong opposites of their nature. In struggling to reconcile these opposites and in striving to achieve creative solutions to the difficult problems which they have set themselves they will often show that psychic turbulence which is so characteristic of the creative person. If, however, we can only recognize the sources of their disturbance, which often enough will result in behavior disturbing to us, we may be in a better position to support and encourage them in their creative striving.

## REFERENCES

Allport, G. W., Vernon, P. E., & Lindzey, G. *Study of values: Manual of directions.* (Rev. ed.) Boston: Houghton Mifflin, 1951.

Barron, F., & Welsh, G. S. Artistic perception as a possible factor in personality style: Its measurement by a figure preference test. *J. Psychol.,* 1952, 33, 199–203.

Bingham, Millicent Todd. Beyond psychology. In *Homo sapiens auduboniensis: A tribute to Walter Van Dyke Bingham.* New York: National Audubon Society, 1953. Pp. 5–29.

Bruner, J. S. *The process of education.* Cambridge, Mass.: Harvard Univer. Press, 1960.

Gough, H. G. *California Psychological Inventory manual.* Palo Alto, Calif.: Consulting Psychologists Press, 1957.

Hall, W. B. The development of a technique for assessing aesthetic predispositions and its application to a sample of professional research scientists. Paper read at Western Psych. Ass., Monterey, Calif., April 1958.

Hathaway, S. R., & McKinley, J. C. *Minnesota Multiphasic Personality Inventory.* New York: Psychological Corporation, 1945.

Jung, C. G. *Psychological types.* New York: Harcourt, Brace, 1923.

Jung, C. G. *Two essays on analytical psychology.* New York: Meridian, 1956.

MacKinnon, D. W. The personality correlates of creativity: A study of American architects. In G. S. Nielsen (Ed.), *Proceedings of the XIV International Congress of Applied Psychology, Copenhagen 1961.* Vol. 2. Copenhagen: Munksgaard, 1962. Pp. 11–39.

Myers, Isabel B. *Some findings with regard to type and manual for Myers-Briggs Type Indicator, Form E.* Swarthmore, Pa.: Author, 1958.

Rexroth, K. My head gets tooken apart. In, *Bird in the bush: Obvious essays.* New York: New Directions Paperbook, 1959. Pp. 65–74.

Russell, W. A., & Jenkins, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8 onr-66216, Office of Naval Research.

Spranger, E. *Types of men.* Trans. P. J. W. Pigors Halle (Saale), Germany: Max Niemeyer, 1928.

Strong, E. K., Jr. *Manual for Strong Vocational Interest Blanks for Men and Women, Revised Blanks (Form M and W).* Palo Alto, Calif.: Consulting Psychologists Press, 1959.

Terman, L. M. *Concept Mastery Test, Form T manual.* New York: Psychological Corporation, 1956.

Welsh, G. S. *Welsh Figure Preference Test: Preliminary manual.* Palo Alto, Calif.: Consulting Psychologists Press, 1959.

Wotton, Henry. *The elements of architecture.* London: John Bill, 1624.

# 25

# Caution in Comparing Creativity and IQ

## — *Richard E. Ripple and Frank B. May*

*The authors report a correlational study on the relationship between intelligence and creative thinking abilities. Groups of junior-high-school students with a wide range of IQs were used as subjects. It was demonstrated that while IQs are not effective predictors of creative-thinking abilities among groups of fairly homogeneous intelligence, to a certain degree they do seem effective in predicting creative-thinking performance in more heterogeneous and representative groups.*

RECENT INVESTIGATIONS (e.g., Getzels & Jackson, 1960; Torrance, 1960) have suggested that there is little relationship between creative-thinking abilities and the generalized ability measured by intelligence tests. Generally, these investigations have been limited to the study of individuals with relatively high IQs.

It is well known that the magnitude of relationship between measures is partially a function of the variability of the groups studied. Limiting the minimum IQ among the Ss studied could lead to mistaken inferences concerning the relationship between intelligence and creative-thinking abilities in groups where intelligence is distributed in a manner more representative of that found in a school classroom.

To demonstrate the caution necessary in expressing the relationship between IQ and creativity, four groups were identified: (a) 30 seventh-grade

Reprinted from *Psychological Reports*, 1962, **10**, 229–230, by permission of the senior author and the publisher.

students (15 girls and 15 boys) whose IQs ranged from 72 to 90 (*Low homogeneous*); (b) 30 seventh-grade students (15 girls and 15 boys) whose IQs ranged from 96 to 110 (*Average homogeneous*); (c) 30 seventh-grade students (15 girls and 15 boys) whose IQs ranged from 116–133 (*High homogeneous*); (d) 30 students (15 girls and 15 boys) identified by selecting 10 students (5 girls and 5 boys) from each of the homogeneous groups. This group was designated *heterogeneous.*

Students in these groups were selected randomly from approximately 1000 students in a large Midwestern city who had been administered the Otis Quick Scoring Mental Ability Test and nine tests of creative thinking. Seven of the creative-thinking tests were adapted from Guilford (1956) for use with children; the other two were devised by May (1961).

Product-moment correlations between IQs and creative-thinking scores were computed separately for each of the four groups. These correlations appear in Table 1. As can be seen, the correlations are considerably smaller in the homogeneous groups than in the heterogeneous group.

TABLE 1. Correlations Between IQ and Creative-Thinking Tests

| | Populations | | | |
|---|---|---|---|---|
| Creative-Thinking Tests | High | Avg. | Low | Het. |
| Expressional Fluency[°] | .39 | .30 | .07 | .73 |
| Plot Questions[°] | .17 | .28 | .22 | .70 |
| Object Improvement[°] | .16 | −.28 | .27 | .30 |
| Metaphors[†] | −.17 | .06 | .07 | .54 |
| Onomatopoeia[†] | .03 | −.06 | .17 | .69 |
| Object Uses-Fluency[°] | .14 | .00 | −.01 | .38 |
| Object Uses-Flexibility[°] | .30 | .20 | .29 | .60 |
| Word Uses-Fluency[°] | −.01 | −.22 | −.01 | .11 |
| Word Uses-Flexibility | .45 | .44 | .12 | .73 |

NOTE. A correlation of .36 is required for significance at the .05 level of confidence; .46 at the .01 level of confidence.

[°] Adapted from Guilford's tests (Guilford, 1956).

[†] Developed by May (1961).

This empirical demonstration of the well-known effect of range of measures on the magnitude of the correlation coefficient emphasizes the need for caution in expressing the relationship between IQs and creative-thinking scores. It appears evident that IQs are not effective predictors of creative-thinking abilities among student populations which are fairly homogeneous with respect to intelligence. However, IQs do seem to be somewhat effective in predicting creative-thinking performance in more representative student populations, that is, student populations which are considerably heterogeneous with respect to intelligence.

*Summary.* By correlating Otis IQs and scores on creative-thinking tests by members of several seventh-grade groups, homogeneous or heterogeneous with respect to IQ, it was demonstrated that the low correlation of these measures reported by other investigators may well be due in part to the restricted IQ ranges in their samples.

## REFERENCES

Getzels, J. W., & Jackson, P. W. The study of giftedness; a multi-dimensional approach. *The gifted student.* Coop. Res. Monogr., No. 2, U.S. Dept. of Health, Education, and Welfare. Washington, D.C.: U.S. Government Printing Office, 1960.

Guilford, J. P., Kettner, N. W., & Christensen, P. R. *A factor-analytic study across the domains of reasoning, creativity, and evaluation:* II. *Administration of tests and analysis of results.* Reports from the Psychological Laboratory: No. 16, Univer. of Southern California, March, 1956.

May, F. B. Creative thinking: a factorial study of seventh-grade children. Unpublished doctoral dissertation, Univer. of Wisconsin, 1961.

Torrance, E. P. Explorations in creative thinking. *Educ.*, 1960, 81, 216–220.

# Part V

## *Learning: Affective Characteristics and Psychomotor Abilities*

AN INQUIRY into the nature of learning would be incomplete without considering the affective characteristics of the learning situation. ("Affect" here refers to the feeling component of human behavior.) The learning of attitudes, values, and other outcomes in the affective domain is necessarily one important objective of the schools. Another objective pertains to the many perceptual motor skills—handwriting, typing, drawing, etc.—that are taught in schools. The efficiency with which these skills are learned is dependent in large measure on our knowledge of the nature of psychomotor abilities. Four selections relevant to affective characteristics and two selections pertaining to psychomotor abilities will be presented then, in Part V as examples of research and thinking in these areas.

The first article, entitled "Attitude Organization in Elementary School Classrooms," is by Jules Henry, professor in the Department of Anthropology at Washington University in St. Louis. In his article, Henry points out that the teacher's private need to "carp and criticize gets in the way of her adequately developing the creative and supportive possibilities in her charges." He recommends promoting teachers' insight into how they project their personal problems into the classroom situation.

Reading number 27 is the report of an investigation relating attitude change to free discussion among groups of different sociometric structures. The investigators, K. M. Miller and J. B. Biggs, were at the University of Tasmania at the time this study was published. (Miller is now with the National Foundation for Educational Research, located in London.) Further research is recommended to ascertain whether or not the attitude changes discovered are a function (a) of the learning process in groups with high cohesiveness, and (b) of personality factors and insecurity in groups with low cohesiveness.

The next selection, by Philip Jackson and Jacob Getzels, is an investigation of the prepotency of psychological variables as compared to environmental variables in understanding dissatisfaction with school. Both Jackson and Getzels are professors at the University of Chicago. One conclusion they come to is that "dissatisfaction with school, like beauty, is frequently in the eye of the beholder."

The self-concept as a psychological construct has enabled educators to better understand children's behavior. Hugh Perkins, professor in the Institute for Child Study at the University of Maryland, reports in reading number 29 a study of factors influencing change in children's self-concepts. On the basis of his findings he hypothesizes that in general, changes in the "self–ideal self" congruency of children may be a manifestation of growth and development, and may reflect the positive influence of school experience.

Turning to psychomotor abilities, a system of such abilities is proposed by J. P. Guilford in reading number 30. Guilford is at the University of Southern California. Using an approach similar to the one used in developing his structure of intellect (see reading number 4), Guilford presents this system as a theory of psychomotor abilities.

Frances Colville, of Santa Barbara College, reports in the final reading an investigation on the effect that knowledge of mechanical principles has on motor-skill performance. Although no evidence supporting a positive effect was found, Colville concludes that some part of the learning period may be devoted to instruction in mechanical principles without detriment to the motor learning of students.

# 26

# Attitude Organization in Elementary School Classrooms — *Jules Henry*

*The author contends that in our culture the elementary-school classroom is one of the most powerful instruments in organizing the attitudes and feelings of youth, so that ultimately through it a social-steady state that is supportive of the culture will be maintained. The classroom teacher is constantly confronted with shifting patterns of feelings and attitudes—her own as well as those of her students. Greater understanding by teachers of their own needs and feelings, and of those of their students may help contribute to the attainment of desired learning outcomes.*

THE WORD *organization* in this paper is used to stand for order and determinateness as distinguished from disorder and randomness. The emotions and attitudes of prepubertal children in our culture are not, on the whole, directed toward generalized social goals, but focused rather on the peer group and parents. From the point of view of an observer who has in mind the larger social goals, like the maintenance of stable economic relations, common front against the enemy, maintenance of positive attitudes toward popular national symbols, and so on, the emotions and attitudes of prepubertal children in our culture may be viewed as lacking order. The adult, on the other hand, is supposed to have so organized his tendencies to

Reprinted from *American Journal of Orthopsychiatry*, 1957, **27**, 117–133, by permission of the author and the American Orthopsychiatric Association, Inc. © The American Orthopsychiatric Association, Inc.

respond to the environment that his emotions, attitudes, and activities subserve over-all social goals. While it is true that attitudes and feelings are bent toward social goals even from earliest infancy,[1] many institutions combine to organize these attitudes and feelings so that ultimately a social steady state will be maintained. The elementary school classroom in our culture is one of the most powerful instruments in this effort, for it does not merely sustain attitudes that have been created in the home, but reinforces some, de-emphasizes others, and makes its own contribution. In this way it prepares the conditions for and contributes toward the ultimate organization of peer- and parent-directed attitudes into a dynamically interrelated attitudinal structure supportive of the culture.

This organizing process is comparable to, though not identical with, the reorganization of attitudes and resources that takes place as a society shifts from a peacetime to a wartime footing. During a period of peace in our society, adult hostility and competitiveness may be aimed at overcoming competition in business or social mobility, while love and cooperation are directed toward family and friends, and toward achieving specific social and economic ends *within* the society. With the coming of war the instruments of government seek to direct hostility and competitiveness toward the enemy, while love and cooperation are directed toward the armed forces, civilian instruments of war (price controls, rationing, civilian officials, etc.), and national symbols. From the point of view of an observer *within the war machine*, the civilian attitudes at first seem random and unorganized. He wants to change them so that from *his point of view* they will seem organized. The situation is similar, though not identical, with respect to the child: to an observer inside the head of even some psychotic children, attitudes and behavior may seem organized. But to the observer on the outside, whose focus is on social goals, the child seems *un-* or *dis*-organized. The prime effort of the adult world is *to make child attitudes look organized to adults*. The emphasis in this paper is on the description of the process of organizing child attitudes as it can be observed in some middle-class urban American classrooms.

## The Witch-Hunt Syndrome

One of the most striking characteristics of American culture since the settlement has been the phenomenon of intragroup aggression, which finds its pathological purity of expression in witch hunts.[2] It comes as a frightening surprise to democratic people to find themselves suddenly in terror of their neighbors; to discover that they are surrounded by persons who carry tales about others while confessing evil of themselves; to perceive a sheeplike docility settling over those whom they considered strong and autonomous. The *witch-hunt syndrome* therefore, as constituting one of the key tragedies of democracy, is selected for the elucidation of the organization of attitudes in our culture. In this witch's brew *destructive criticism* of others is the toad's horns, *docility* the body of the worm; *feelings of vulnerability* the chicken heart; *fear of internal (intragroup) hostility* the snake's fang; *confession of evil deeds* the locust's leg; and *boredom and emptiness* the dead man's eye. The witch-hunt syndrome is thus stated to be a dynamically interrelated system of feelings and actions made up of destructive criticism of others, docility, feelings of vulnerability, fear of internal aggression, confession of evil deeds, and boredom.

---

[1] J. Henry and J. W. Boggs, "Child Rearing, Culture, and the Natural World," *Psychiatry*, 1952, **15**, pp. 261–271.

[2] M. L. Starkey, *The Devil in Massachusetts* (New York: Alfred A. Knopf, Inc., 1949).

The witch-hunt syndrome in full panoply was observed in but one of the dozen classrooms in four schools studied in the research which I discuss here. Thus it seems a relatively rare phenomenon. But the question I set myself to answer is, How could it occur at all? What are the attitudes, present in the children, that were organized by this teacher into the syndrome? How could she do it? With what materials did she works? She did not create out of nothing the attitudes she manipulated in her "Vigilance Club" in this fourth-grade classroom in a middle-class American community. She had to have something to start with. The argument of this paper will be that the feelings and tendencies to action which this teacher organized into the witch-hunt syndrome in her class are present in an *un*organized state in other classrooms. Given a certain type of teacher, he or she will be able to develop into a highly specialized, tightly integrated system in his classroom those attitudes which are present in differently organized state in the children in all classrooms. Let us now look at a meeting of the Vigilance Club.[3]

1. In the extreme back of the room is a desk called the "isolation ward." A child has been placed there for disclipinary reasons. The Vigilance Club of the class is holding a meeting. . . . Officers are elected by the group. The purpose of the club is to teach children to be better citizens. The order of procedure is as follows: the president . . . bangs her gavel on the desk and . . . says, "The meeting of the Vigilance Club will come to order." Each child then takes from his or her desk a booklet whose title is *All About Me* . . . and places it on top of his desk. The vice-president calls the name of a child, gets the child's booklet, and places it on the teacher's desk. The president then calls on the child and asks, "——, have you been a good citizen this week?" The president says, "Name some of the good things you have done," and the child tries to recall some, like opening doors for people, running errands, etc. Next the president asks the class if it remembers

---

[3] In order to prevent identification of teachers and children, the names of my student observers are not used.

any good things the child has done. Each point is written in the child's booklet by the teacher. The president then . . . says to the child, "Name the bad things you have done. . . ." The child reports the wrongs he has committed during the week, and the class is asked to contribute information about his behavior. This too is written in the booklet by the teacher, who also reprimands the student, registers horror, scolds, etc. . . . When one child reports a misdemeanor of another the teacher asks for witnesses, and numerous children sometimes volunteer. . . . The child in the "isolation ward" reported some good deeds he had done; the children reported some more, and the isolated child was told he would soon be released. . . . [During this meeting some children showed obvious pleasure in confessing undesirable behavior. One child, by volunteering only good things of the students, seemed to be using the situation to overcome what seemed to the observer to be her unpopularity with the class.]

Before analyzing this protocol for the attitudes present in it, it will be well to look at some events that occurred in this classroom on another day.

2. During the game of "spelling baseball" a child raised her hand and reported that Alice and John had been talking to each other. This occurred when neither child was "at bat." The teacher asked Alice if this was so, and she replied that it was, but John denied having spoken to Alice. The teacher said that John must have listened to Alice, but he denied this too. Then the teacher asked whether there had been any witnesses, and many hands were raised. Some witnesses were seated on the far side of the room, and hence could not have seen Alice and John from their location in the room. All those testifying had "seen" Alice talking, but denied John's guilt. Alice was sent to the "bull pen," which meant that she had to sit on the floor behind the teacher's desk, and could no longer participate in the game. . . .

3. Mary raised her hand and said, "It hurts me to say this. I really wish I didn't have to do it, but I saw Linda talking." Linda was Mary's own teammate, had just spelled a word correctly, and had gone to first base. The teacher asked Linda if she had talked, and Linda said, "No, I just drew something in the air with my finger. . . ." She was sent to the "bull pen."

In these examples we see intragroup aggression; docility of the children in conforming, with no murmur of protest, to the teacher's wishes; and confession of "evil." In such a situation children develop feelings of vulnerability and fear of detection. Let us now look for these phenomena in classrooms presided over by teachers who seem to represent the more normal American type, in comfortable, middle-class, white communities: teachers who are conscientious and reasonably gentle, but creatures of their culture, and humanly weak. We begin not with internal aggression as expressed in spying and talebearing, but with the milder, though closely releated phenomenon of carping, destructive criticism. While this occurs throughout the sample, I give here examples only from a fifth-grade classroom in the same school system.

4. Bill has given a report on tarantulas. As usual the teacher waits for volunteers to comment on the child's report.

Mike: The talk was well illustrated, well prepared. . . .

Bob: Bill had a piece of paper [for his notes], and teacher said he should have them on cards. . . .

Bill says he could not get any cards.

Teacher says that he should tear the paper next time if he has no cards.

Bob: He held the paper behind him. If he had had to look at it, it wouldn't have looked very nice.

5. Betty reports on Theodore Roosevelt.

A child comments that it was very good but she looked at her notes too much.

Teacher remarks that Betty had so *much* information.

Bob: She said "calvary" [instead of "cavalry"].

6. Charlie reads a story he made up himself: "The Unknown Guest." One dark, dreary night . . . on a hill a house stood. This house was forbidden territory for Bill and Joe, but they were going in anyway. The door creaked, squealed, slammed. A voice warned them to go home. Spider webs, dirty furniture . . . Bill wanted to go home. They went upstairs. A stair cracked. They entered a room. A voice said they might as well stay and find out now; and their father came out. He laughed and they laughed, but they never forgot their adventure together.

Teacher: Are there any words that give you the mood of the story? . . .

Lucy: He could have made the sentences a little better. . . .

Teacher: Let's come back to Lucy's comment. What about his sentences?

Gert: They were too short. . . .

Charlie and Jeanne are having a discussion about the position of the word "stood."

Teacher: Wait a minute, some people are forgetting their manners. . . .

Jeff: About the room: the boys went up the stairs and one "cracked"; then they were in the room. Did they fall through the stairs or what?

Teacher suggests Charlie make that a little clearer.

Lucy: If he fell through the step. . . .

Teacher: We still haven't decided about the short sentences. Perhaps they make the story more spooky and mysterious.

Gwynne: I wish he had read with more expression instead of all at one time.

Rachel: Not enough expression.

Teacher: Charlie, they want a little more expression from you. I guess we've given you enough suggestions for one time. (Charlie does not raise his head, which is bent over his desk as if studying a paper.) Charlie! I guess we've given you enough suggestions for one time, Charlie, haven't we? (Charlie half raises his head, seems to assent grudgingly.)

The striking thing about these examples is that the teacher supports the children in their carping criticism of their fellows. Her performance in this is not, however, consistent; but even where, as in Example 6, she seems at one point to try to set herself against the tide of destruction, by calling attention to the possible artistry in Charlie's short sentences, she ends up supporting the class against him, and Charlie becomes upset. Thus teacher, by rewarding the children's tendencies to carp, reinforces them. Teachers, however, are able to make their own contributions to this tendency. The single example given below will serve as illustration:

7. Joan reads us a poem she has written about Helen Keller . . . which concludes with the couplet:

> "Helen Keller as a woman was very great;
> She is really a credit to the United States."

Teacher (amusedly): Is "states" supposed to rhyme with "great"? When Joan murmurs that it is, the teacher says, "We'll call it poetic license."

From time to time one can see a teacher vigorously oppose tendencies in the children to tear each other to pieces. The following example is from the sixth grade:

8. The Parent-Teachers Association is sponsoring a school frolic, and the children have been asked to write jingles for the publicity. For many of the children the experience of writing a jingle seems painful. They are restless, bite their pencils, squirm around in their seats, speak to their neighbors, and from time to time pop up with questions like, "Does it have to rhyme, Mr. Smith?" . . . At last Mr. Smith says, "All right, let's read some of the jingles now." Child after child says he "couldn't get one"; but some have succeeded. One girl has written a very lone jingle *obviously the best in the class*. However, instead of using Friday as the frolic day she used Tuesday, and several protests were heard from the children. Mr. Smith defended her. "Well, so she made a mistake. But you are too prone to criticize. If *you* could only do so well!"

It will be observed that all the examples are taken from circumstances in which the child's self-system is most intensively involved; where his own poetry or prose is in question, or where he has worked hard to synthesize material into a report. It is precisely at the points where the ego is most exposed that the attack is most telling. The numerous instances in the sample, where the teachers, by a word of praise or a pat on the head, play a supportive role, indicate their awareness of the vulnerability of the children. Meanwhile, as I have pointed out, the teachers often fall into the trap of triggering or supporting destructive impulses in the children.

The carping criticism of one's peers is a form of intragroup aggression, which can be quite threatening and destructive. Talebearing, however, countenanced by some teachers more than by others, can be an overwhelming threat to autonomy. While telling on others can be organized into the patrol-monitor complex (prestige through controlling and telling), useful perhaps in maintaining order in large school populations, its operation within the classroom may have serious consequences. Let us look at a couple of examples:

9. Second grade. As teacher asked the children to clear their desks one boy raised his hand, and when called on said, "Jimmy just walked by and socked me on the head."

Teacher: Is this true?

Jimmy: He hit me first.

Teacher: Why don't you both take seats up here (in front of the room). I'm not sure people like you belong in the second grade.

10. Sixth-grade special class for bright students.

The children are working on their special nature study projects. Ralph passes where Joseph is working. Ralph (to teacher): Joseph is writing too much on his birds.

Teacher: Joseph, you should write only a few things.

In our sample, telling on other children in the classroom is infrequent outside the class in which the Vigilance Club was formed. Destructive criticism is the preferred mode of attack in most other classrooms. The ease with which tendencies to attack peers can be organized into telling on others, however, is illustrated by the monitor-patrol complex, and by the Vigilance Club (Example 3).

## Competition

Competition is an important element in the witch-hunt syndrome. Since witch hunts involve so often obtaining the attention and approval of some powerful central figure, the examples of competitiveness that I shall cite illustrate how approval and attention seeking occur as the child attempts to beat out his peers for the nod of the teacher. It would be easy to cite examples from the protocols of the merciless laughter of children at the failures or gaucheries of their classmates. I am interested, however, more in showing the all-pervading character of the phenomenon of competition, *even in its mildest forms*. The first example is from a fourth-grade music lesson:

11. The children are singing songs of Ireland and her neighbors from the book *Songs of Many Lands*. . . . Teacher plays on piano while children sing. . . . While children are singing some of them hunt in the index, find a song belonging to one of the four countries, and raise their hands before the previous song is finished in order that they may be called on to name the next song. . . .

Here singing is subordinated, in the child, to the competitive wish to have the song he has hunted up in the index chosen by the teacher. It is merely a question of who gets to the next song in the index first, gets his hand up fast, and is called on by the teacher.

The following examples also illustrate the fact that almost any situation set by the teacher can be the occasion for release of competitive inmpulses:

12. The observer enters the fifth-grade classroom.

Teacher: Which one of you nice polite boys would like to take [observer's] coat and hang it up? (Observer notes: From the waving hands it would seem that all would like to claim the title.)

Teacher chooses one child . . . who takes observer's coat. . . .

Teacher: Now children, who will tell [observer] what we have been doing?

Usual forest of hands . . . and a girl is chosen to tell. . . .

Teacher conducted the arithmetic lesson mostly by asking, "Who would like to tell . . . the answer to the next problem?"

This question was usually followed by the appearance of a large and *agitated* forest of hands; apparently *much competition to answer.*

Thus the teacher is a powerful agent in reinforcing competition.

It has already been pointed out that carping criticism helps to settle in the child a feeling of vulnerability and threat. In this connection it is significant that *the failure of one child is repeatedly the occasion for the success of another*. I give one illustration below from the same class as the one from which I have taken Example 12.

13. Boris had trouble reducing 12/16 to lowest terms, and could get only as far as 6/8. Much excitement. Teacher asked him quietly [note how basically decent this teacher is] if that was as far as he could reduce it. She suggested he "think." Much heaving up and down from the other children, all frantic to correct him. Boris pretty unhappy. Teacher, patient, quiet, ignoring others, and concentrating with look and voice on Boris. She says, "Is there a bigger number than 2 you can divide into the two parts of the fraction?" After a minute or two she becomes more urgent. No response from Boris. She then turns to the class and says, "Well, who can tell Boris what the number is?" Forest of hands. Teacher calls, Peggy. Peggy gives 4 to be divided into 12/16, numerator and denominator.

Where Boris has failed Peggy has been triumphant; *Boris's failure has made it possible for Peggy to succeed.*

This example and also Example 6 are ones in which the discomfort of the child was *visible,* and such instances may be multiplied. They illustrate how vulnerable the children feel in the presence of the attacks of the peer group in the classroom. But since these are children who face the world with serious anxiety to begin with, the classroom situation sustains it. Let us look at some stories created by these very children, and read by them to their classmates. We have already seen one, Example 6, Charlie's story of "The Unknown Guest." Here are *all* the stories read to their classmates by these children during an observation period.

14. (a) Charlotte's story: "Mistaken Identity." One day last year my family

and I went to the hospital to visit somebody. When we were coming out and were walking along my father hit me. I came up behind him to hit him back, but just as I was about to do it I looked back and he was behind me! I was going to hit the wrong person!

(b) Tommy's story: "The Day Our House Was Robbed." [Observer has recorded this in the third person.] He was coming home from school one afternoon. He knew his Mom was away that afternoon. He started to go in the side door, but decided, he doesn't know why, to go round the back. He found the door open, went into the kitchen, looked into the front room where he saw a thief. Tommy "froze stiff" (chuckle of appreciation from the class), ran out, shouted "Stop thief" as the man ran out after him. He went to a neighbor, rang the bell, called his mother at the store. The cops came, asked questions, but the man had gotten away with $99 and his mother's watch. If he had gone in the side door he would not have had a chance to see the man. Changing to the back door "may have saved my life." [Teacher's only remarks about this story were: 1) instead of having said "froze stiff," Tommy should have said "froze stiff as something"; 2) he should have left out the word "then" in one place; 3) he could have made the story clearer; 4) he changed from the past to the present tense.]

(c) Polly's story: "Custard the Lion." Custard the Lion was the most timid animal in Animal Town. The doctors couldn't cure him. Then they found a new medicine. It had strange effects, but Custard wanted to try it. When he did he felt very queer. (Child gives details of queer feeling.) But he soon realized he wasn't afraid of anything. [Teacher's first remark: "You didn't let us hear the last sentence."]

(d) Dan's story: "The Boy Hero." Bill wanted to be a fireman, so he went to the firehouse. The Chief was telling him to go home when the bell clanged. While the Chief was getting into the engine, he didn't see that Bill was getting on too. (Class or teacher picks up flaw in sentence and it is reread correctly.) The Chief said O.K. as long as Bill was aboard, "But you're not to get into no mischief." (Class choruses, "Any. . . .") Everyone was out of the fire except a little girl and her doll. The firemen cannot figure out what to do, but Bill, seeing a tree near the house, climbs it over the protests of the firemen. He misses the girl on his first try, but gets her on the second. While sliding down the tree she slips and almost falls, but grabs Bill's pants, and they make it to safety. . . . [children's remarks center on position of "clang, clang, clang" in the story. Teacher talks about how to use direct quotations, which, it seems, Dan had not used properly.]

(e) Bertha's story: [Title not recorded.] The story is about Jim who was walking home past the Smith's house one night and heard a scream. Penny Smith came out and said there was a robber in the house. When the cops came they found a parrot flying around in there, and Penny's parents told her to shut the parrot up before she read mystery stories again. [This story was followed by much carping criticism, which was terminated by the teacher's telling Bertha to change the story to suit the class.]

These stories contain elements of anxiety and even of terror. As each child finishes, the carping criticism of students and teacher then reminds him of his vulnerability. As the child sends out his cloud of fear, it returns with the leaden rain of hostility.

### Docility

It comes as a somewhat shocking surprise, perhaps, to middle-class parents, to find their children described as "docile." Yet we have already seen the perfection of docility in the Vigilance Club, and we shall presently see its manifold forms in more normal classrooms.

15. First grade. The children are to act out a story called "Pig Brother," which is about an untidy boy. The teacher is telling the story. One boy said he did not like the story, so the teacher said he could leave if he did not wish to hear it again, but the boy did not leave.

16. In gym the children began to tumble, but there was much restless activity in the lines, so the teacher had all the children run around the room until they were somewhat exhausted before she continued the tumbling.

17. Second grade. The children have been shown movies of birds. The first film ended with a picture of a baby bluebird.

Teacher: Did the last bird ever look as if he would be blue?

The children did not seem to understand the "slant" of the question, and answered somewhat hesitantly, yes.

Teacher: I think he looked more like a robin, didn't he?

Children, in chorus: Yes.

Item 17 is one of a large number of instances, distributed throughout all grades, in which the children exhibit their docility largely through giving the teacher what he wants. Thus in the elementary schools of the middle class the children get an intensive eight-year-long training in hunting for the right signals and giving the teacher the response wanted. The rest of the examples of docility document this assertion.

18. Fourth grade. (a) An art lesson.

Teacher holds up a picture.

Teacher: Isn't Bob getting a nice effect of moss and trees?

Ecstatic Ohs and Ahs from the children. . . .

The art lesson is over.

Teacher: How many enjoyed this?

Many hands go up.

Teacher: How many learned something?

Quite a number of hands come down.

Teacher: How many will do better next time?

Many hands go up.

(b) Children have just finished reading the story "The Sun Moon and Stars Clock."

Teacher: What was the highest point of interest—the climax?

The children tell what they think it is. Teacher is aiming to get from them what *she* considers the point of climax, but the children seem to give everything else but.

Bobby: When they capture the thieves.

Teacher: How many agree with Bobby?

Hands, hands.

19. Fifth grade. This is a lesson on "healthy thoughts," for which the children have a special book depicting, with appropriate illustrations, specific conflictual incidents among children. The teacher is supposed to discuss each incident with the children in order to help them understand how to handle their emotions.

One of the pictures is as follows: A sibling *pair* is illustrated by *three* boys: 1) One has received a ball. 2) One is imagined to react with displeasure. 3) One is imagined to react benignly and philosophically, saying, "My brother couldn't help being given the football; we'll use it together."

Teacher: Do you believe it's easier to deal with your thoughts if you own up to them, Betty?

Betty: Yes it is, if you're not cross and angry.

Teacher: Have you any experience like this in the book, Alice?

Alice tells how her brother was given a watch and she envied him and wanted one too; but her mother said she wasn't to have one until she was fifteen, but now she has one anyway.

Teacher: How could you have helped—could you have changed your thinking? How could you have handled it? What could you do with mean feelings?

Alice seems stymied. Hems and haws.

Teacher: What did Susie (a character in the book) do?

Alice: She talked to her mother.

Teacher: If you talk to someone you often then feel that "it was foolish of me to feel that way. . . ."

Tommy:—He had an experience like that, he says. His cousin was given a bike and he envied it. But he wasn't "ugly" about it. He asked if he might ride it, and his cousin let him, and then, "I got one myself; and I wasn't mean, or ugly or jealous."

Before continuing it will be well to note that since the teacher does not say Alice was wrong the children assume she was right and so copy her answer.

Two boys, the dialogue team, now come to the front of the class and dramatize the football incident.

Teacher (to the class): Which boy do you think handled the problem in a better way?"

Rupert: Billy did, because he didn't get angry. . . . It was better to play together than to do nothing with the football.

Teacher: That's a good answer, Rupert. Has anything similar happened to you, Joan?

Joan can think of nothing.

Sylvester: I had an experience. My brother got a hat with his initials on it because he belongs to a fraternity, and I wanted one like it and couldn't have one; and his was too big for me to wear, and it ended up that I asked him if he could get me some letters with my initials, and he did.

Betty: My girl friend got a bike that was 26-inch, and mine was only 24; and I asked my sister what I should do. Then my girl friend came over and was real nice about it, and let me ride it.

Teacher approves of this, and says, Didn't it end up that they both had fun without unhappiness?

Here we note that the teacher herself has gone astray, for on the one hand her aim is to get instances from the children in which they have been yielding, and capable of resolving their own jealousy, etc.; yet, in the instance given by Betty, it was not Betty who yielded, but her friend. The child immediately following Betty imitated her since Betty had been praised by the teacher.

Matilde: My girl friend got a 26-inch bike and mine is only 24; but she only let me ride it once a month. But for my birthday my mother's getting me a new one, probably (proudly) a 28. (Many children rush in with the information that 28 doesn't exist.) Matilde replies that she'll probably have to raise the seat then, for she's too big for a 26.

As we go on with this lesson, we shall continue to see how the children's need for substitute gratification and their inability to accept frustration are the real issues, which even prevent them from getting the teacher's point. We shall see how, in spite of the teacher's driving insistence on her point, the children continue to inject their conflicts into the lesson, while at the same time they gropingly try to find a way to gratify the teacher. *They* cannot give the "right" answers because of their conflicts; teacher cannot handle their conflicts, even perceive them, because *her* underlying need is to be gratified by the children. The lesson goes on:

Teacher: I notice that some of you are only happy when you get your own way. You're not thinking this through, and I want you to. Think of an experience when you didn't get what you want. Think it through.

Charlie:—His ma was going to the movies and he wanted to go with her, and she wouldn't let him; and she went off to the movies, and he was mad; but then he went outside and there were some kids playing baseball, so he played baseball.

Teacher: But suppose you hadn't gotten to play baseball? You would have felt hurt, because you didn't get what you wanted. We can't help feeling hurt when we are disappointed. What could you have done; how could you have handled it?

Charlie: So I can't go to the movies, so I can't play baseball, so I'll do something around the house.

Teacher: Now you're beginning to think! It takes courage to take disappointments. (Turning to the class) What did we learn? The helpful way . . .

Class: is the healthy way!

Before entering the final section of this paper, we need to ask: Why are these children, whose fantasies contain so many hostile elements, so docile in the classroom; and why do they struggle so hard to gratify the teacher and try in so many ways to bring themselves to her attention (the "forest of hands")? We might, of course, start with the idea of the teacher as a parent figure, and the children as siblings competing for the teacher's favor. We could refer to the unresolved dependency needs of children of this age, which make them seek support in the teacher, who manipulates this seeking and their sibling rivalry to pit the children against each other. Other important factors, however, that are inherent in the classroom situation itself, and particularly in middle-class classrooms, ought to be taken into consideration. We have observed the children's tendency to destructively criticize each other, and the teachers' often unwitting repeated reinforcement of this tendency. We have taken note of the anxiety in the children as illustrated by the stories they tell, and observed that these very stories are subjected to a carping criticism, whose ultimate consequence would be anything but alleviation of that anxiety. Hence the classroom is a place in which the child's underlying anxiety may be heightened. In an effort to alleviate this he seeks the approval of the teacher, by giving right answers and by doing what teacher wants him to do under most circumstances. Finally, we cannot omit the teacher's need to be gratified by the attention-hungry behavior of the children.

A word is necessary about these classrooms as middle class. The novel *Blackboard Jungle*[4] describes schoolroom behavior of lower-class children. There we see the children *against the teacher*, as representative of the middle class. But in the classes I have described we see the *children against each other*, with the teacher abetting the process. Thus, as the teacher in the middle-class schools directs the hostility of the children toward one another and away from himself, he reinforces the competitive dynamics within the middle class itself. The teacher in lower-class schools, on the other hand, appears to become the organizing stimulus for behavior that integrates the lower class, as the children unite in expressing their hostility to the teacher.

## Confession

The Vigilance Club would have been impossible without confession, and the children's pleasure in confession. But, as with the other parts of the syndrome, confessing occurs in other classrooms also; it can be elicited when the proper conditions are present, and the children can be seen to enjoy it —to vie with one another in confessing. Let us follow the lesson on "healthy thoughts" a little further. We will see how confession occurs as the children seek to give teacher *precisely* what she wants.

---

[4] E. Hunter, *The Blackboard Jungle* (New York: Simon and Schuster, 1954).

20. Teacher asks if anyone else has had experience like that [of two children who have just recited], where they were mean and angry.

Dick:—He has a friend he plays baseball with, and sometimes they fight; but they get together again in a few minutes and apologize.

In this first example we note one of the important aspects of the confession element in the syndrome: the culprit must have given up his evil ways, and now be free of impurities.

In response to Dick's story, teacher says: You handled it just right. Now let's hear about someone who had a similar experience and didn't handle it just right.

Tom:—His little brother asked for the loan of his knife, but it was lost, and he got angry with his little brother for asking. [This knife story follows a sequence of several stories about knives told by other children. The exuberance of knife stories following immediately on the teacher's approval of the first one suggests that some of them are made to order and served up piping hot for teacher's gratification.]

Teacher: Now Tom, could you have worked it out any differently? (Observer notes that Tom seems to enjoy this confession; certainly he is not abashed or ashamed.)

Tom: Later he asked me if he could help me find it. He found it in a wastebasket, and then I let him borrow it.

Harry: Sometimes I get angry when my friends are waiting for me and . . . (observer missed some of this) and my little sister asked if she could borrow my autoracing set, and I hit her once or twice. (Class laughs.)

Here we see another factor so important to the flourishing of the syndrome: the audience gets pleasure through the confessor's telling about deeds the audience wishes to commit: who among Harry's listeners would not like to have hit his sister, or anyone, "once or twice"?

The teacher then goes on: What would you do now—would you hit her?

Harry: Now I'd probably get mad at first, but let her have it later.

Thus Harry has mended his ways—in teacher-directed fantasy at least—and returned to the fold.

So far we have had confession of mean and angry thoughts and violence. We shall now see confession to unacceptable fear. In all cases the teacher says what type of confession she wishes to hear, and what the resolution should be of the unacceptable behavior; and the children vie with one another to tell commensurable tales, as they derive pleasure from the total situation—through approval of the teacher, expression of their own real or fantasied deviations, and the delight of their peers. In these situations the pleasure of the peer group is seen to derive not so much from the "happy ending" the children give their stories but rather from the content of the story itself. It is interesting that no carping criticism appears; rather the entire situation is a jolly one. It seems that within unspoken limits the children permit one another to boast of "evil" behavior because of the deep pleasure obtained from hearing it. Thus impulse expression becomes a device for role maintenance in the classroom.

The lesson proceeds:

> Two children enact a little skit in which they have to go to the principal to ask him something. One of them is afraid of the principal, the other is not. The moral is that the princpal is the children's friend, and that one should not be shy.
>
> Gertrude: Well, anyway, the principal isn't a lion, he's your friend; he's not going to kill you.
>
> Teacher: That's right, the principal is a friend, he says hello and good morning to you. . . . Have you ever felt shy?
>
> Meriam: The first year I sold Girl Scout cookies I didn't know how to approach people; and the first house I went to I didn't know the lady; and I stuttered and stammered, and didn't sell any cookies. By the second house I had thought it all out before I rang the bell, and I sold two boxes. (Triumphantly.)
>
> Teacher: It helps to have self-confidence.
>
> Ben now tells a story, with a happy ending, of being afraid of a principal. Then Paul tells a story, amid gales of laughter about his being scared on a roller coaster. By this time there is so much excitement among the children that the teacher says: Wait a minute—manners!
>
> John:—He was scared to go on the Whip-the-Whirl (scornful laughter from the class); but after he went he liked it so much that he went eight times in a row. (This is well received.)
>
> Many hands go up. Teacher waits. . . .
>
> Michael:—He was at Pleasure Park on the ferris wheel (scornful Aw from the class) "and a girl kept rocking it, and I started to get green" (roar of laughter).
>
> Teacher: Now we'll have to stop.

Certain phenomenon not emphasized before appear in this section. Confession is used by the authoritative figure, the teacher, to strengthen attachment to significant but potentially terrifying figures like school principals, and to polish up cultural shibboleths like "self-confidence." For the child storytellers confession becomes an opportunity for bathing in the emotional currents of the peer group, as the child stimulates the group's approval through presentation of group standards, and awakens group pleasure as the peer group responds to its own anxiety about weakness, and experiences resolution of the anxiety through the happy ending. With a perfect instinct for what is right, each child provides catharsis for his peers. By presenting himself as weak, he enables his peers to identify with him; and then, as he overcomes his weakness, he enables his companions, too, to feel strong.

What this lesson on healthy thoughts may have accomplished by way of creating a permanent reservoir of "healthy thoughts" is difficult to say, but that it helped create solidarity among the students, and between them and the teacher is clear from the fact that when she suddenly shifted ground to say, "Do you think you are wide enough awake for a contest in subtraction of fractions?" the children responded with a unanimous roar of "Yes," as if she had asked them whether they were ready for cookies and ice cream!

Thus in this lesson, in which all have participated more with their *unconscious* than with their conscious emotions, solidarity has been achieved. Teacher thought she was teaching the children to have healthy thoughts, but she was showing them how to gratify her. The children sensed this and struggled to gratify her, while they sought acceptance by their peers also. The essential difference between this teacher and the one who perpetrated the Vigilance Club is that though the latter tended to demolish solidarity among the children while placing the teacher in supreme command, the lesson on healthy thoughts tended to a dubious solidarity among all. *Both teachers organize some of the same elements in the children, but into different configurations, of total feeling and behavior.*

## Boredom

It seems unnecessary to document the fact that children become bored in class, for much of modern thinking and curriculum arrangement is aimed at eliminating it. The shifts at 15-minute intervals from one subject to the next in the elementary school classrooms is one example of this effort. Boredom, which means emotional and intellectual separation from the environment, is an insupportable agony, particularly if the emotional vacuum created by such separation is not filled by gratifying fantasies, or if it is filled by terrifying ones. To fill this vacuum people in our culture will throw themselves into a great variety of even relatively ungratifying activities. Since in this situation, bored children attack almost any novel classroom activity with initial vigor, the witch-hunt syndrome or any modification thereof helps to overcome boredom: better to hunt than be bored. In a full and satisfying life there is no place for witch hunts. The school system that can provide a rich program for children has no need of Vigilance Clubs, nor even of lessons on "healthy thoughts."

## Discussion and Conclusions

In this paper I have used suggestions from communications theory in an effort to order the data obtained from direct observation of elementary school classrooms. Information, the central concept of communications theory, refers to measurable differences in states of organization. In human behavior, as seen in the classroom under discussion, we observe *qualitative shifts in state, for different teachers organize the same underlying emotional characteristics of the children to achieve different organizations of the emotions.* One teacher so organizes the children's emotions as to accomplish an intensification of the fear of intragroup aggression, while she turns the children's hostility toward one another. A different teacher may organize the emotions of the children so that a euphoria in which students and teacher are bathed in a wave of emotional gratification is achieved. The great skill in being a teacher would seem to be, therefore, a *learned*-capacity to keep shifting states of order intelligently as the work demands. This does not mean the traditional classroom order, where you can hear a pin drop, but rather the kind of order in which the *emotions of the children are caught up and organized toward the achievement of a specific goal.* It is not necessary, perhaps, that even the most prominent emotions of the children, like competitiveness, for example, form part of the organized whole. Yet, on the other hand, it is difficult to see how, in the present state of our culture, competitiveness can be overlooked. It would seem, perhaps, that the important outcome to avoid is that the competitiveness should become destructive of peers, while reinforcing dependence on the teacher.

The phenomenon I have labeled "docility" occurs because of the absolute dependence for survival of the children on the teacher. That is to say success in school depends absolutely on the teacher, and self-respect, as a function of the opinion of others, in the home or among peers, is in part a function of success or failure in school. In these circumstances the child's capacity to respond automatically to the signals he gets from the teacher is bound to acquire somewhat the appearance of instinctive behavior. Although it occurs at a much higher level of integration than instinct, the child hunts for the proper signals from the teacher, and the child's responses take on instinctual quality. They *must;* otherwise, like the nestling who does not open its mouth when the mother arrives with a worm, he will never eat the ambrosia of teacher's approval, so necessary to his survival. In this situation both children and teacher easily become the instruments of their own unconscious processes, as they, like Joseph and his brethren, fall on each other's necks in a shared ecstasy of exuberant dependence. Teacher and pupil will have gratified each other, but it remains an open question whether the children will have learned what the curriculum committee planned.

We see in the organization of the components of the witch-hunt syndrome an important phase in the formation of American national character, for tendencies to docility, competitiveness, confession, intragroup aggression, and feelings of vulnerability the children may bring with them to school, are reinforced in the classroom. This means that independence and courage to challenge are observably played *down* in these classrooms. It means, on the other hand, that tendencies to own up rather than to conceal are reinforced—a development which, in proper hands, might become a useful educational instrument. It means, further, that while many teachers do stress helping others they may inadvertently develop in the children the precise opposite, and thus undermine children's feelings of security. One could come from a very secure and accepting family and yet have one's feelings of security and acceptance threatened in these classrooms. On the other hand, what seems most in evidence from the stories they make up is that the children come to school with feelings of vulnerability which are intensified in the classroom.

Meanwhile we should try to understand that all the teachers in the sample were probably trying to be good teachers,[5] and all the children were trying to be good pupils. Their unconscious needs, however, naturally dominated their behavior. The teacher who organized the Vigilance Club probably thought she was teaching her children to be upright and honest, and to perform good deeds, but her unconscious tendencies caused these worthy inclinations to seek the wrong expression. All teachers need conformity in the classroom in order that the children shall absorb a respectable amount of academic knowledge. But the teacher's (often unconscious) need for acceptance by the children, and her fear (sometimes unconscious) of her inability to control free discussion, compel her to push the children into uncritical docility at times, while they seek her approval.

The creation of stories, and their discussion by the class, are accepted principles of progressive education. But the teacher's own (at times unconscious) need to carp and criticize gets in the way of her adequately developing the creative and supportive possibilities in her charges. Thus these are not "bad," "vicious," or "stupid" teachers, but human beings, who express in their classroom behavior the very weaknesses parents display in their dealings with their children. The solution to the problem of the contradiction between the requirements of a democratic education on the one hand, and the teachers' unconscious needs on the other, is not to carp at teachers, and thus repeat the schoolroom process, but to give them some insight into how they project their personal problems into the classroom situation.

---

[5] I am indebted to B. Bettelheim for this suggestion.

# 27

# Attitude Change Through Undirected Group Discussion — *K. M. Miller and J. B. Biggs*

*This study, conducted in England, is concerned with the relationship between measured attitude changes among adolescent boys toward various racial groups, and undirected group discussion and group structure. Boys were assigned, on the basis of sociometric choices, to groups of high and low cohesiveness for the purpose of discussing freely and without any outside guidance a number of racial groups. A control group was identified but did not participate in the discussions. No difference in tolerance between the discussion groups resulted; however, they did show increased tolerance over the control group. Implications for attempts to change attitudes are discussed in terms of group characteristics.*

STUDIES OF attitude change have emphasized many variables; the present study was concerned with the effectiveness of free group discussion about racial groups when the discussion groups are sociometrically structured. Some writers, (*3, 8, 11*) have stressed the importance of using sociometric structure as an aid to effective classroom work, suggesting that the educational process is more efficient when groups are composed of mutually attracted members, i.e., when groups are cohesive.

Reprinted from *Journal of Educational Psychology*, 1958, 49, 224–228, by permission of the senior author and the American Psychological Association.

## Design of Study

Two types of groups were selected from a class—psychegroups considered high, and sociogroups low in cohesion. Attitude change was assessed by testing with an attitude scale before and after a period of undirected discussion about a number of racial groups. The stability of any change was assessed by a third test some weeks later. A control class completed the attitude scale at the same times but without intervening discussion.

The interval between the first and second tests was designed to minimize memory of responses to the first. School vacation prevented the interval between the second and third tests being identical with the first interval. On no occasion were Ss informed that subsequent tests would be given.

## Techniques

*Sociometric.* The conventional form of the Moreno technique was used, Ss being asked to write the names (up to five for each category) of those classmates next to whom they would like to sit and would not like to sit.

*Social Attitudes.* A Bogardus type scale was selected as the most suitable both for repeated measurement and for showing changes after discussion. The form used was similar to the Zeligs and Hendrickson (13) modification but had been independently derived by the senior author for a previous study. The steps were:

I would like to have live in my home.
I would like to have as a close friend.
I would like to go for a holiday with.
I would like to have in my sports team.
I would like to work with in school.
I would like to have live in my street.
I would like to have live in my country.

So the list of racial groups would be meaningful for the Ss, 14 were selected either on the basis of percentage of national group among migrants to Australia or for historical reasons, e.g., English and Japanese. The groups were: American, Chinese, Dutch, English, German, Indian, Irish, Italian, Japanese, Jewish, Negro, Polish, Russian, Balt.

Scoring was by the Zeligs method (12) whereby each positive response counted one point.

### Subjects

Two third year secondary school classes of 26 and 16 boys respectively were selected. Of the larger class 24 members with a mean age of 177 months, *SD* 8.5, were experimental Ss while all of the smaller class with a mean age of 180 months, *SD* 9.0 months, were control Ss. The difference in age was not significant.

### Selection of Discussion Groups

The method of analysis of sociometric data suggested by Clark and Maguire (2) was used, supplemented in the final choice of groups by reference to sociograms. Three friendship or psychegroups and three neutral or sociogroups each containing four boys were selected. The psychegroups were chosen so that within each (i) all members expressed strong (1, 2, 3) choices for each member of the group (ii) each member received at least one mutual choice and (iii) no one was rejected by other members of the group. The composition of the sociogroups was such that no member had expressed either acceptance or rejection of any other member.

### Procedure

The task was presented as part of a general study being conducted in several countries, to find out how children thought about people in their own and other countries.[1] To encourage frankness Ss were assured that all replies were confidential and would not be seen by anyone in the school. The sociometric and social distance scales were then administered to both control and experimental classes.

The initial administration of the social distance scale was as recommended by Bogardus (*1*), the *E* reading the items at three-second intervals. On the later occasions Ss were allowed to complete the scale at their own rate.

Four weeks later the group discussions were begun, friendship and neutral groups working alternately. Each group was instructed only after assembling and at the conclusion of the discussion the members were asked not to discuss the activity with the other boys. The *E* introduced the discussion, explaining that he would like each member to say something about a number of racial groups. The discussion period lasted approximately 30 minutes, two minutes being allowed for each of the 14 races, though discussion of any one group was not abruptly terminated.

---

[1] A report is in preparation on the International study, the design of which is described in (*7*).

During the discussion, *E* was passive and nondirective, averting any questions put directly to him. Following the discussion the social distance scale was readministered. After half of the group discussions had been completed the scale was given a second time to the control group.

As a check on the stability of any change the scale was given a third time two weeks after the last group discussion to both control and experimental classes.

## Results and Analysis

*Comparison of control and experimental groups.* A *t* test of initial scores revealed no significant difference ($t = 1.10$, 38 *df*), thus showing that the mean attitude level of the two could be considered equivalent.

*The effect of discussion.* The mean scores for each of the three sections—psychegroups, sociogroups, and control class—on first and second administrations were tested for differences. The differences for the friendly and neutral Ss were significant beyond the one per cent level ($t = 3.43$ and 3.16, 11 *df respectively*) while the difference for the controls was not significant ($t = 0.32$, 15 *df*).

*Differences between second and third administrations.* Inspection of Table 1 shows that the mean scores for the third administration are lower than immediately after the discussion period in the case of the friendly and neutral Ss. These differences were not significant for any of the three sections (*t* values of 2.14, 1.27, and 0.86 for friends, neutrals, and controls respectively).

TABLE 1. Mean Social Distance Scores for Friend and Neutral Ss by Group

| | Administration | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Friends | | | |
| Group 1 | 33.25 | 47.25 | 42.25 |
| Group 2 | 49.25 | 72.50 | 57.50 |
| Group 3 | 52.00 | 54.25 | 55.75 |
| Total | 43.75 | 58.00 | 51.83 |
| Neutrals | | | |
| Group 1 | 39.50 | 51.25 | 50.50 |
| Group 2 | 37.75 | 70.00 | 63.00 |
| Group 3 | 51.00 | 57.75 | 54.30 |
| Total | 42.75 | 59.67 | 56.00 |

*Differences between first and third administrations.* A similar analysis was made of differences between first and third sets of scores. The *t* test analysis showed that final scores were significantly greater than the initial scores for both psychegroup and sociogroup Ss; at the two per cent level for the former and at the five per cent level for the latter. Again the differences in the control class scores were not significant ($t = 0.33$).

*Quantitative aspects of the discussions.* The remarks made by each S were recorded and categorized as favorable, unfavorable or neutral towards the race under discussion. The total remarks are shown in Table 2 where it is seen that the total number of remarks and the distribution according to category are approximately equal for the psychegroups and sociogroups. Examination showed that in the psychegroups remarks were somewhat more evenly spread over all members than in the sociogroups.

TABLE 2. Number of Remarks of Each Type Made During Discussion

| | Favorable | Unfavourable | Neutral | Total |
|---|---|---|---|---|
| Psychegroups | 151 | 133 | 42 | 326 |
| Sociogroups | 155 | 129 | 39 | 323 |

## Discussion

The analysis has shown that the members of both psychegroups and sociogroups show more tolerance (decreased social distance) after free undirected discussion of some characteristics of different racial groups. No such change is shown by control Ss tested after the same interval of time. The amount of initial change appears to be unrelated to type of group as the mean difference between friends and neutrals is not significant.

When the same scale is applied again after an interval of two weeks the mean scores of members of both types of experimental group are closer to the mean scores in the first testing than they were on the second. The amount of this reversion is, however, not statistically significant, indicating that the favorable change engendered by the discussion was fairly stable over this short period. Further confirmation was provided by the comparison of the initial and final scores which were significantly different for both psychegroups and sociogroup subjects, at two and five per cent levels respectively.

Few, if any, investigators have suggested that free undirected discussion about racial groups would lead to the measurable changes demonstrated in this study. Some investigators (6, 9) indicate that attempts to change attitudes are more successful when Ss are members of naturally working groups. While a class is in some respects a functioning group it has within it a number of groups which are more cohesive than the class as a whole. Thus it would not have been surprising to find the members of the psychegroups showing greater and more stable change than the sociogroup members. The results of the present study are not in accord with such expectations as both friends and neutrals show (approximately) equally significant changes and stability of change. Moreover, they did not differ from each other in degree of change throughout.[2]

The present findings are, however, in keeping with the suggestion that close friendships may lead to better communication and wider participation in the group (*6*). The evidence for better communication is provided by notes of the actual discussion sessions. The climate in the psychegroups was freer and more lively and discussion more spontaneous than that in the sociogroups where discussion tended to be more formal and reserved, though no less frequent, and not different in the proportion of favorable, unfavorable and neutral remarks. Thus although quantiative differences between friend and neutral groups were not discovered, there is a difference in the way in which the quantitative result was achieved.

Some investigators (*8, 10*) have considered such changes in terms of group conformity, suggesting that there would be a greater tendency for members of sociogroups, being less cohesive than psychegroups, to establish some norm or common position. Others (*4, 9*) have considered such changes as a function of security–insecurity and personality adjustment–maladjustment suggesting that less secure, less well-adjusted persons may, as a means of establishing a more secure group relationship, change towards a central position. When the results of the present study were examined for evidence of conformity it was found that for all three sociogroups and for none of the psychegroups the range of scores after discussion was considerably smaller than before.

The relevance of these findings for education requires consideration as the results seem to be at variance with those usually claimed by proponents of the sociometric approach. Work of investigators such as Cunningham, Oeser, and Shoobs suggests that group discussion would be more effective in the psychegroups than in the sociogroups, whereas it has been shown in this study that measured changes are approximately equal for both types of groups. Further research is required to ascertain whether the changes are primarily a function of the learning process, in one case, and a function of personality factors and insecurity in the other.

## Summary and Conclusions

This study was an attempt to relate attitude change with free discussion in psyche- and sociogroups. Several findings are definite while others merit further investigation.

---

[2] Unrelated $t$ tests on differences scores between psychegroup and sociogroup members at each stage were nonsignificant.

1. Free, undirected discussion about racial groups by two types of small groups, selected on a sociometric basis, resulted in a significant change of attitude irrespective of the type of group. Further, this change was relatively stable over a short period.

2. Contrary to expectations from sociometric studies in the classroom, and from studies of group structure, no significant differences between the quantitative changes of friendly and neutral Ss were discovered.

3. Nevertheless, it was suggested that the psychological processes in the two types of groups might be different and that further investigation is necessary to show whether the tendency for the scores of members of sociogroups to come closer to a central position after discussion is a function of personality adjustment.

## REFERENCES

1. Bogardus, E. S. Measurement of personal group relations. *Sociometry*, 1947, **10**, 306–311.
2. Clark, R. A., & McGuire, G. Sociographic analysis of sociometric valuations. *Child Develpm.*, 1952, **23**, 141–154.
3. Cunningham, Ruth, et al. *Understanding group behavior of boys and girls.* New York: Teachers College, Columbia Univer., 1951.
4. Grossman, B., & Wrighter, J. Relationship between selection-rejection and intelligence, social status, and personality amongst 6th grade children. *Sociometry*, 1948, **11**, 346–355.
5. Kelley, M., & Thibaut, J. Experimental studies in group problem solving and process. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison Wesley, 1954. P. 762.
6. Lewin, K., & Grabbe, P. Conduct, knowledge, and acceptance of new values. *J. soc. Issues*, 1945, **1**, 53–74.
7. Miller, K. M. Evaluation in adult education. *Internat. soc. sci. Bull.*, 1955, **7**, 430–442.
8. Oeser, O. A. *Teacher, pupil and task.* London: Tavistock, 1955.
9. Sherif, M., & Sherif, C. W. *Groups in harmony and tension.* New York: Harper, 1953.
10. Sherif, M., & Sherif, C. W. *Outline of social psychology.* (Rev. ed.) New York: Harper, 1956. Pp. 301–328.
11. Shoobs, N. E. Sociometry in the classroom. *Sociometry*, 1947, **10**, 233–241.
12. Zeligs, Rose. Children's intergroup attitudes. *J. genet. Psychol.*, 1948, **72**, 101–110.
13. Zeligs, R., & Hendrickson, G. Racial attitudes of 200 6th grade children. *Sociol. soc. Res.*, **18**, 26–36.

# 28

# Psychological Health and Classroom Functioning: A Study of Dissatisfaction with School Among Adolescents

### — *Philip W. Jackson and Jacob W. Getzels*

*In considering reasons why certain students are dissatisfied with school, a suggestion often made is that they are unable to cope adequately with scholastic materials. Students at the other end of the scholastic scale, however, who are of high ability, may be dissatisfied because they experience a lack of stimulation in their classrooms. Still another, and third source of dissatisfaction relates to more basic personality variables, ones that are less dependent on academic success or boredom. In the present article, the authors have investigated the relationship between these variables and school dissatisfaction. Their findings indicate a greater relevance than had before been suggested of psychological health data, and also the importance of differentiating the attitudes of girls from boys, in understanding dissatisfaction with school.*

THE PROBLEM of dissatisfaction with school among children is of theoretical and practical significance to both psychologists and educators. At the theoretical level dissatisfaction with school becomes part of a broader area of inquiry which aims at an understanding of the individual's functioning in an institutional setting and which includes studies of staff morale,

Reprinted from *Journal of Educational Psychology*, 1959, **50**, 295–300, by permission of the senior author and the American Psychological Association.

role conflict, productivity, and the like. At a practical level the question of why children like or dislike school is directly related to the immediate problems of school dropouts, grouping procedures, planning for the gifted child, and the like.

As might be expected, a social phenomenon as important as dissatisfaction with school is not without its explanatory hypothesis. Some of these spring from empirical findings, while others appear to be part of our cultural ethos. Educational studies that point to an empirical linkage between school failure and school dropouts, and industrial studies that demonstrate a relationship between low morale and decreased output, lead one to suspect that reduced effectiveness in school (i.e., low scholastic achievement) would be a natural concomitant of dissatisfaction with the institution. Thus one would expect to find heightened dissatisfaction among students who have low ability or who are unable for one reason or another to deal adequately with scholastic material.

More recently it has been suggested (although never adequately demonstrated) that many successful students with high ability are dissatisfied with their school experiences; the term "boredom" is often linked with the term "gifted child" in current expositions by educators. The boredom problem among "gifted" combined with the failure experiences of the low ability child suggests that the greatest number of dissatisfied students is to be found among extreme ability groups. Those who are low in ability and achievement would be expected to show dissatisfaction because of the numerous frustrations they experience in the classroom. Those who are high in ability and achievement would be expected to show dissatisfaction because of the relative lack of stimulation which they experience in the classroom.

Both of these explanations (or, more accurately, hypotheses) contain the implication that dissatisfaction with an institution arises out of the individual's interaction with that institution. An alternative explanation might be that the individual brings a set toward satisfaction or dissatisfaction *to* the institution—that it is a reflection of a more pervasive personal orientation and that success or failure experiences within the institution have a limited influence upon it. This hypothesis obviously places more emphasis than do the earlier ones upon psychological variables, as opposed to environmental variables, in understanding dissatisfaction with school. The research described here was designed to test the relative merit of these alternative views.

## Problem

The purpose of this investigation is to examine the differences in psychological functioning and classroom effectiveness between two groups of adolescents—those who are satisfied with their recent school experiences and those who are dissatisfied.

## Subjects and Procedure

The Ss of this investigation were two groups of adolescents identified from among 531 students enrolled in a Midwestern private school. These students were divided into five class groups ranging from the prefreshmen to the senior year of high school. In this institution a single grade, the prefreshmen, is substituted for the usual seventh and eighth grades. The instrument used to select the experimental groups, called the Student Opinion Poll, was a 60-item opinionnaire designed to elicit responses concerning general satisfaction or dissatisfaction with various aspects of school—viz., the teachers, the curriculum, the student body, and classroom procedures. The following are sample items, one in each of the four areas.

3. While there are some differences among them, most teachers in this school are:
   - *a.* Very inspiring
   - *b.* Quite inspiring
   - *c.* Somewhat inspiring
   - *d.* Not inspiring

16. Most of the subjects taught in the school are:
   - *a.* Interesting and challenging
   - *b.* Somewhat above average in interest
   - *c.* Somewhat below average in interest
   - *d.* Dull and routine

14. From the standpoint of intellectual ability, students in this school are:
   - *a.* Too bright—it is difficult to keep up with them
   - *b.* Just bright enough
   - *c.* Not bright enough—they do not provide enough intellectual stimulation

5. The freedom to contribute something in class without being called upon by the teacher is:
   - *a.* Discouraged more than it should be—students do not have enough opportunity to have their say
   - *b.* Encouraged more than it should be—students seem to be rewarded just for speaking even when they have little to say
   - *c.* Handled about right

The instrument was scored by giving one point each time the S chose the "most satisfied" response to a multiple-choice item. Thus, the possible range of scores was from 0 to 60. For the total school population the mean score on the Student Opinion Poll was 37.30; the standard deviation was 9.57. The experimental groups were chosen as follows:

Group I—the "dissatisfied" group—consisted of all students whose score on the opinionnaire was at least one and a half standard deviations *below* the mean of the entire student body. This group contained 27 boys and 20 girls.

Group II—the "satisfied" group—consisted of all students whose score on the opinionnaire was at least one and a half standard deviations *above* the mean of the entire student body. This group contained 25 boys and 20 girls.

The experimental groups were compared on the following variables:

1. *Individual intelligence tests.* In most cases this was the Binet. A small number of children were given the Henmon-Nelson, the scores of which were converted by regression equation into equivalent Binet scores.

2. *Standardized verbal achievement test.* The Cooperative Reading Test was used. Prefreshmen and freshmen were given Test $C_1$, Form Y; older students were given $C_2$, Form T.

3. *Standardized numerical achievement tests.* Because of differences in the curricula of the various grade groups it was not possible to administer the same test of numerical achievement to all Ss. The following tests were given according to grade placement:

Prefreshman—Iowa Everypupil Arithmetic Test, Advanced Form O.
Freshmen—Snader General Mathematics Test.
Sophomores—Cooperative Elementary Algebra Test, Form T.
Juniors—Cooperative Intermediate Algebra Test.
Seniors—Cooperative Geometry Test, Form 2.

4. *California Personality Test.* Two forms of this instrument were used. The intermediate form was given to prefreshmen; the secondary form was given to all of the older groups. Two subscores were obtained, "personal adjustment" and "social adjustment."

5. *Direct Sentence Completion Test.* Ss were asked to complete 27 sentences of the type: "When I saw I was going to fail I . . . ," or, "I think my father is . . . ." Each sentence was given a plus or minus score depending upon the presence or absence of morbid fantasy, defeatism, overt aggression, and the like. The total score was the summation of the individual sentence scores.

6. *Indirect Sentence Completion Test.* This instrument was identical with the Direct Sentence Completion Test except that proper names were inserted for the pronoun "I," thus changing it from a "self-report" to a "projective" instrument. Boys' names were used in the male form of the instrument and girls' names in the female form. The instrument was presented as a "thinking speed" test. To reinforce this notion Ss were asked to raise their hands when they were finished and the elapsed time was written on their test booklet. This instrument was administered approximately two weeks prior to the administration of the Direct Sentence Completion Test.

7. *Group Rorschach.* Cards III, IV, IX, and X were projected on a screen. For each picture the S was presented with 10 responses and was asked to choose the three which he thought to be most appropriate. Each list of 10 contained four "pathological" responses. The S's score was the number of nonpathologic responses among his 12 choices. This group technique follows that described by Harrower-Erikson and Steiner (1945).

8. *Teacher ratings.* Each student was given three ratings by his present teachers. These ratings included: (*a*) his general desirability as a student; (*b*) his ability to become involved in learning activities; and (*c*) his possession of leadership qualities. Teachers were required to place all of their students on a five-point scale so that Categories 1 and 5 each contained one-twelfth of the students; Categories 2 and 4 each contained one-fourth of the students; and Category 3 contained one-third of the students. The values 5, 8, 10, 12, and 15 were assigned to the categories and were used in quantifying the ratings.

9. *Adjective Check List.* From a list of 24 adjectives each student was asked to choose the 6 which best described his characteristic feelings while attending classes in particular school subjects. The list contained 12 "positive" (e.g., confident, happy, eager, relaxed) and 12 "negative" adjectives (e.g., bored, restless, misunderstood, angry). The use of the negative adjectives by the experimental groups was analyzed both quantitatively and qualitatively.

## Results

With the exception of the adjective check list the results of all comparisons are shown in Table 1. Contrary to popular expectations the "satisfied" and "dissatisfied" students did *not* differ from each other in either general intellectual ability or in scholastic achievement. Those differences which did appear were linked to psychological rather than scholastic variables. More specifically, each of the test instruments designed to assess psychological health or "adjustment" was effective in distinguishing "satisfied" from "dissatisfied" students within one or both sex groups.

For both sexes the experimental groups were differentiated by their scores on the California Test of Personality. The experimental groups of boys were further differentiated by their responses to the Indirect Sentence Completion Test. For girls additional differences appeared in their responses to the Direct Sentence Completion Test and the Group Rorschach.

On all of these test variables the "satisfied" group attained the "better" score—i.e., the score signifying a more adequate level of psychological functioning. It is also worthy of note that whenever a significant difference appeared, the mean score of the total student population fell between the

TABLE 1. Mean Scores, Standard Deviations, and *t* Statistics for Satisfied and Dissatisfied Adolescents on Dependent Variables[a]

| | Boys | | | | | Girls | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Dissatisfied ($N=27$) | | Satisfied ($N=25$) | | *t* | Dissatisfied ($N=20$) | | Satisfied ($N=20$) | | |
| | $\bar{x}$ | *s* | $\bar{x}$ | s | | $\bar{x}$ | *s* | $\bar{x}$ | *s* | *t* |
| IQ | 134.85 | 14.58 | 136.44 | 14.59 | ns | 128.45 | 15.06 | 128.00 | 11.45 | ns |
| Verbal Achievement | 49.96 | 8.69 | 50.68 | 7.87 | ns | 50.63 | 9.11 | 52.28 | 6.76 | ns |
| Numerical Achievement | 50.35 | 9.75 | 52.17 | 10.52 | ns | 47.78 | 8.61 | 48.50 | 10.26 | ns |
| Calif. Personal Adjust. | 45.58 | 9.82 | 53.40 | 7.63 | 3.18°° | 47.90 | 13.03 | 54.76 | 9.25 | 1.86° |
| Calif. Social Adjust. | 44.85 | 11.37 | 51.84 | 8.93 | 2.45°° | 47.00 | 13.15 | 55.76 | 7.89 | 2.50°° |
| Direct Sentence Comp. | 46.93 | 10.58 | 49.25 | 10.02 | ns | 46.65 | 12.01 | 54.00 | 5.73 | 2.53°° |
| Indirect Sentence Comp. | 47.19 | 9.61 | 51.29 | 6.95 | 1.75° | 49.60 | 10.35 | 53.47 | 7.97 | ns |
| Group Rorschach | 48.35 | 10.66 | 47.44 | 10.30 | ns | 47.35 | 11.35 | 54.16 | 8.32 | 2.15°° |
| Teacher Rating I: Desirability as a student | 8.94 | 1.83 | 10.35 | 1.70 | 2.85°° | 9.84 | 1.91 | 10.05 | 1.59 | ns |
| Teacher Rating II: Leadership qualities | 9.01 | 2.08 | 10.13 | 1.96 | 2.00° | 9.91 | 2.37 | 10.04 | 1.24 | ns |
| Teacher Rating III: Involvement in learning | 9.09 | 2.14 | 10.23 | 1.69 | 2.14°° | 9.67 | 2.32 | 10.33 | 2.11 | ns |

° Significant at the .05 level.
°° Significant at the .01 level.

[a] With the exception of IQ, all scores were based upon parameters of the total student body from which the experimental groups were drawn. The scores of all tests were transformed to T scores with a mean of 50 and a standard deviation of 10. For the total population the teacher ratings have a mean of 10 and a standard deviation of 2. The mean IQs for the total school population are: boys, 132, and girls, 128.

mean scores of the experimental groups. Thus, the variables that differentiate the experimental groups tend also to distinguish them from the total population of students.

In addition to showing differences on psychological health variables, "satisfied" and "dissatisfied" boys were perceived differently by their teachers. On all three of the teachers' ratings the "satisfied" boys received more favorable judgments than did "dissatisfied" boys. The fact that this result does not appear to be true for girls lends support to the popular expectation that boys are more likely to express their negative feelings publicly than are girls. This hypothesis receives some confirmation from the results of the adjective check list which are described below.

In Table 2 are shown the number of Ss who chose negative adjectives when asked to describe their typical classroom feelings. As they are arranged in Table 2 the adjectives reflect the rankings of four judges who were asked to rank the words on the degree to which they involved an implicit or explicit

Table 2. Number of Subjects Choosing Negative Adjectives when Asked to Describe Typical Classroom Feelings

| | Boys | | | Girls | | |
|---|---|---|---|---|---|---|
| Adjective | Dissatisfied ($N = 27$) | Satisfied ($N = 25$) | Chi Square | Dissatisfied ($N = 20$) | Satisfied ($N = 20$) | Chi Square |
| Inadequate | 19 | 16 | ns | 17 | 7 | 10.42°° |
| Ignorant | 19 | 13 | ns | 15 | 3 | 14.54°° |
| Dull | 25 | 16 | 6.36° | 16 | 9 | 5.60° |
| Bored | 24 | 13 | 8.61°° | 20 | 13 | 8.48°° |
| Restless | 20 | 15 | ns | 19 | 9 | 11.90°° |
| Uncertain | 20 | 21 | ns | 17 | 13 | ns |
| Angry | 15 | 4 | 8.76°° | 13 | 4 | 8.29°° |
| Unnoticed | 19 | 5 | 13.25°° | 7 | 4 | ns |
| Unhelped | 18 | 8 | 6.24° | 9 | 6 | ns |
| Misunderstood | 16 | 5 | 8.31°° | 5 | 2 | ns |
| Rejected | 12 | 3 | 6.66°° | 4 | 0 | ns |
| Restrained | 17 | 2 | 16.91°° | 9 | 3 | 4.29° |

° Significant at the .05 level.
°° Significant at the .01 level.

criticism of others. The 12 adjectives were typed on separate cards and were accompanied by the following directions:

> On the following cards are a number of negative adjectives which a person might use to describe himself. Rank these adjectives on the degree to which they involve an implicit or explicit criticism of others. For each adjective ask the question: If a person used this adjective *to describe himself* would he also be implicitly or explicitly criticizing others? Give a rank of 1 to the adjective which would be *least* critical of others and a rank of 12 to the adjective which would be *most* critical of others.

Four psychologists served as judges. The average rank order correlation among the four sets of judgments was .84. The adjectives are presented in Table 2 according to the ranked sum-of-ranks of the judges. The adjective "inadequate" was judged as being most free of criticism of others, while the adjective "restrained" was judged as involving the greatest amount of criticism of others.

As might be expected, the use of negative adjectives was far more frequent among dissatisfied students than among satisfied students. Four adjectives seemed to discriminate equally well between the experimental groups for both sexes; these were: "bored," "angry," "restrained," and "dull."

An examination of Table 2 also suggests the existence of sex differences in the students' description of their typical classroom feelings. Remembering the classificatory scheme by which the adjectives are ranked in Table 2, it appears that dissatisfied girls are somewhat less likely than dissatisfied boys to use negative adjectives involving implicit criticism of others. Dissatisfied boys, on the other hand, are less likely than dissatisfied girls to be distinguished from their satisfied counterparts by the use of adjectives *not* involving implicit criticism of others. If one thinks of criticism directed towards others within Rosenzweig's schema of "intropunitiveness" and "extrapunitiveness" (Murray, 1938), then the observed sex differences may be conceptualized by saying that dissatisfied girls are more *intropunitive* than satisfied girls; dissatisfied boys are more *extrapunitive* than satisfied boys.

This difference in the direction of aggression may provide a context for the obtained differences in teacher ratings discussed earlier. If the dissatisfied boy is more likely than his female counterpart to lay the blame for his dissatisfaction upon others in his environment, particularly school authorities, it is reasonable to expect that he would be viewed as somewhat less than completely desirable by the classroom teacher. The dissatisfied girl, on the other hand, seems more willing to direct her negative feelings inward, thus avoiding the additional risk of counter-aggression by school authorities or by other adults.

## Discussion

Two major conclusions are suggested by the findings of this study. First, dissatisfaction with school appears to be part of a larger picture of psychological discontent rather than a direct reflection of inefficient functioning in the classroom. It is almost as if dissatisfaction were a product of a pervasive perceptual set that colors the student's view of himself and his world. Second, it appears that the "dynamics" of dissatisfaction operate differently for boys and girls. Boys seem to project the causes of their discontent upon the world around them so that adults are seen as rejecting and lacking in understanding. This tendency to blame adults may be one reason why these boys are seen as less attractive by teachers than are satisfied boys. Girls, on the other hand, are more likely to be self-critical, turning blame for their dissatisfaction inward. Feelings of inadequacy, ignorance, and restlessness more sharply differentiate satisfied and dissatisfied girls than is the case with boys. This tendency to be intropunitive may partially explain why teacher ratings fail to distinguish between our two experimental groups of girls.

The atypicality of the sample population used in this research places a number of limitations upon the inferential statements which can be made on the basis of these findings. Fortunately, however, the major portion of the investigation has recently been replicated using seventh and eighth grade lower-class Negro adolescents as Ss (Spillman, 1959). The findings of the latter study are essentially the same as those reported here. Again the psychological rather than the intellectual or scholastic variables discriminated between satisfied and dissatisfied students. The findings with respect to the use of negative adjectives were not as clear-cut but, again, every intropunitive adjective was used more frequently by dissatisfied girls as compared with dissatisfied boys, while the latter exceeded the girls in their use of extrapunitive adjectives.

It should be noted that even the most satisfied students made some use of negative adjectives when asked to describe their typical feelings in the classroom. Also, the average member of the satisfied group expressed some dissatisfaction on one-sixth of the questions in the Student Opinion Poll. These two observations should serve as ample cautions against the danger of interpreting any sign of dissatisfaction with school as symptomatic of deeper psychological difficulties. Apparently, some degree of dissatisfaction is the rule rather than the exception. Nonetheless, the responses of the extremely disgruntled group of students leaves little doubt that dissatisfaction with school, like beauty, is frequently in the eye of the beholder.

## Summary

This investigation examines the differences in psychological functioning and classroom effectiveness between two groups of adolescents—those who are satisfied with their recent school experiences and those who are dissatisfied. The major findings point to: (*a*) the relevance of psychological health data rather than scholastic achievement data in understanding dissatisfaction with school; (*b*) the importance of differentiating the attitudes of dissatisfied girls from those of dissatisfied boys, the former being characterized by feelings of personal inadequacy, the latter by feelings critical of school authorities. Rosenzweig's concepts of intropunitiveness and extrapunitiveness are applied to these findings and a relevant theoretical framework is proposed.

## REFERENCES

Harrower-Erikson, M. R., & Steiner, M. E. *Large scale Rorschach techniques.* Springfield, Ill.: Charles C. Thomas, 1945.

Murray, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.

Spillman, R. J. Psychological and scholastic correlates of dissatisfaction with school among adolescents. Unpublished master's thesis, Univer. of Chicago, 1959.

# 29

# Factors Influencing Change in Children's Self-Concepts — *Hugh V. Perkins*

*A major purpose of education is to promote individual self-development. The writer of this article suggests that study of the self-concept phenomenon can help achieve this purpose. Self-concept is defined as those perceptions, beliefs, feelings, attitudes, and values which the individual views as describing himself; whereas the ideal self is defined as those qualities which describe the person as he would like to be. The author reports an investigation of changes in the congruence of self-concept and ideal self among fourth- and sixth-graders. He goes on to relate the degree of this congruence with the differential effects of certain independent variables (e.g., sex, grade level, social-emotional climate).*

IN A PREVIOUS article (5) this writer suggested that the self-concept phenomenon can serve education in two ways: (a) as a psychological construct which enables teachers, counselors, parents, and others to achieve with training deeper understandings and insights into the behavior and development of children, and (b) as a vital and important aspect of learning and development which the school through its educational processes seeks to promote and foster in every child. The need for a further examination and understanding of the uses of this construct in education provided the focus for an investigation made by this writer of selected factors influencing perceptions of and changes in children's self-concepts. A previous article

Reprinted from *Child Development*, 1958, **29**, 221–230, by permission of the author and the Society for Research in Child Development, Inc.

reported the methodology and the findings of this research which related to teachers' and peer groups' perceptions of children's self-concepts (5). This article will report the findings relating to the influence of selected factors on changes in children's self-concepts.

This second part of the study investigated the effects of four factors, (a) social-emotional climate, (b) teacher participation in an in-service child study program, (c) teacher acceptance of self and others, and (d) grade level, on changes in children's self-concepts in the fourth and sixth grades.

Self-concept was defined as those perceptions, beliefs, feelings, attitudes, and values which the individual views as describing himself. The ideal self was defined in terms of those qualities which describe the person he would like to be. In this study data on children's self-concepts and ideal selves were obtained by means of a Q-sort, an instrument and technique which requires the subject to make a forced normal distribution of a series of self-referrent statements. The Q-sort developed for use in this study was a stratified random sample of 50 self-referrent statements which had been selected from a universe derived from responses of fourth and sixth grade children reported in a study by Jersild (3). Correlating the child's self-sort with his ideal sort provides a measure of his self-ideal self congruency and may be expressed in terms of a Q-correlation coefficient. Change in self-concept in this study was defined as change in self-ideal congruency.

A representative sample of eight classrooms enrolling 251 children in seven suburban elementary schools of a large county school system in Maryland was selected for study. Classrooms were selected on the basis of their being willing to participate in the research and their being representative of one of the extremes with respect to each of the independent variables: (a) social-emotional climate, group-centered *vs.* teacher-centered, (b) participation in child study, completion of the three year program in child study *vs.* no participation in program, (c) teacher acceptance of self and others, more acceptance *vs.* less acceptance of self and others, and (d) grade level, fourth *vs.* sixth.

Classroom climates, group-centered and teacher-centered, were identified using Withall's Climate Index (7). Participation in child study referred to teacher participation for three years in the in-service child study program sponsored by the Institute for Child Study of the University of Maryland. Teacher scores on Reed's Sentence Completion Test (6) were used as measures of teacher acceptance of self and others. Scores on the California Test of Mental Maturity, 1953 S Form, and the California Achievement Tests in Language, Reading, and Arithmetic, Primary and Intermediate Forms A and B, were used in ascertaining the relationships between reliability in using the Q-sort and reading and mental age and in correlating changes in school achievement with changes in self–ideal self congruency.

The general hypothesis tested by the second part of this study was as follows: Children in group-centered climates, those whose teachers have participated in child study, those whose teachers have a greater acceptance of self and others, and those in sixth grade will show greater congruency between their self-concepts and ideal selves than will children, respectively, in teacher-centered climates, those whose teachers have never participated in child study, those whose teachers have less acceptance of self and others, and those in fourth grade.

In addition, the following specific hypotheses were tested:

1. Children generally will register positive and significant changes through time toward increased congruency between their self-concepts and ideal selves.
2. Girls will register a significantly greater congruency between their self-concepts and ideal selves than will boys.
3. Children who register greater change in school achievement and peer group acceptance will also show greater change toward congruency between their self-concepts and ideal selves.

## Findings

### Reliability of Q-sort Instrument

The results of the reliability test of children in performing the Q-sort, though not a major problem of this investigation, are reported because of the clues they suggest regarding the stability and consistency of children's self-concepts. This test of reliability of children's use of the Q-sort was accomplished by asking the children in each classroom to make a repeated self-sort (self-concept) two to seven days after they had made an initial self-sort. This test of reliability was made during the second collection of data after the children had become familiar with the instruments. Mean *z* scores and corresponding reliability coefficients for repeated administration of the Q-sorts are reported in Table 1 by sex, classroom, and grade.

The reliability coefficient of .65 for the Q-sort, though somewhat lower than had been anticipated, indicates, nevertheless, that the Q-sort is a reliable instrument for measuring the self-concepts of children in groups. Since the reliability test of an instrument is also a test of the reliability of persons using the instrument, the observed reliability coefficients of fourth and sixth grade children's use of the Q-sort in this study is indicative of the level of consistency of their self-concepts at these maturity levels. Thus, the relative consistency of children's self-concepts related not only to the reliability of the Q-sort instrument, but it was also an important aspect of the phenomenon studied by the research.

TABLE 1. Mean *z* Scores and Corresponding Reliability Coefficients Obtained from Repeated Administrations of Q-Sort

| Teacher and class | Boys *N* | Boys Mean *z* | Boys *r* | Girls *N* | Girls Mean *z* | Girls *r* | Entire Class *N* | Entire Class Mean *z* | Entire Class Corresponding *r* |
|---|---|---|---|---|---|---|---|---|---|
| Fourth Grade: | | | | | | | | | |
| A | 18 | .616 | .55 | 20 | .736 | .63 | 38 | .679 | .59 |
| B | 9 | .801 | .66 | 6 | .766 | .64 | 15 | .787 | .66 |
| J | 18 | .418 | .40 | 18 | .785 | .66 | 36 | .601 | .54 |
| K | 16 | .651 | .51 | 18 | .760 | .64 | 34 | .666 | .58 |
| Entire Fourth | 61 | .599 | .54 | 62 | .762 | .64 | 123 | .683 | .59 |
| Sixth Grade: | | | | | | | | | |
| C | 13 | .802 | .66 | 18 | .924 | .73 | 31 | .876 | .70 |
| D | 20 | .816 | .67 | 18 | 1.028 | .77 | 38 | .916 | .72 |
| L | 13 | .970 | .75 | 10 | .935 | .73 | 23 | .955 | .74 |
| M | 13 | .674 | .59 | 23 | .810 | .67 | 36 | .761 | .64 |
| Entire Sixth | 59 | .816 | .67 | 69 | .924 | .73 | 128 | .877 | .70 |
| Entire Sample | 120 | .707 | .61 | 131 | .843 | .69 | 251 | .780 | .65 |

Other results of the test for reliability reported in Table 1 indicate that girls' reliability in using the Q-sort instrument was .69 whereas the boys' was .61. Moreover, sixth grade children, whose coefficient was .70, appeared to be more reliable than fourth graders whose coefficient was .59. These differences in reliability between boys and girls and between fourth and sixth grade children were analyzed for significant differences using the normal curve sampling distribution. For a two-tailed test of significance, neither the *z* value of 1.06 for sex differences nor the *z* value of 1.51 for grade level differences was significant. Since the *z* value of 1.51 approaches significance at the 5 per cent point for a one-tailed test, the hypothesis that more mature children revealed greater consistency in their self-concepts than do less mature children should be tested by further research.

In this research it was assumed that children in fourth and sixth grades have sufficient insight into their selves and sufficient skills in reading and understanding to enable them to use the Q-sort instrument. This assumption suggests that children's reliability in using the Q-sort may be related to reading level and mental age. To ascertain whether such relationships exist, product-moment correlation coefficients were obtained by correlating *z* scores of reliability coefficients with grade level scores in reading and with mental age. These are reported in Table 2.

The results reported in Table 2 indicate generally that relationships between Q-sort reliability and reading level and mental age are positive but low. These findings are corroborated by those of Brandt (1) who found a low positive relationship between accuracy of self-estimate and intelligence.

TABLE 2. Product-Moment Correlation Coefficients by Classrooms Obtained by Correlating Q-Sort Reliability $z$ Scores with Reading Levels and Mental Ages

| Teacher and Classroom | Grade | Q-Sort Reliability $z$ Score *vs.* Reading Level | Q-Sort Reliability $z$ Score *vs.* Mental Age |
|---|---|---|---|
| A | 4 | .57 | .52 |
| B | 4 | .12 | .43 |
| J | 4 | .64 | .36 |
| K | 4 | .14 | .08 |
| C | 6 | .26 | .27 |
| D | 6 | .51 | .51 |
| L | 6 | .20 | .08 |
| M | 6 | .46 | .43 |
| Corresponding $r$ of Mean $z$ | | .38 | .34 |

In the present study it will be noted that correlations of Q-sort reliability with reading level and mental age are quite similar for most classrooms. This suggests that reading performance correlates highly with mental age since Q-sort reliability operated as a constant in these comparisons. From these results it is concluded that other factors such as personal adjustment are probably more closely related to consistency of self-concept than are reading level and mental age.

## Changes in Children's Self-Concepts

Change in self-concept has been defined as change in congruency of the child's self-concept and his ideal self occurring during a specified period of time. An individual may be presumed to be moving toward better adjustment and to be facilitating his self-development when his self-concept and ideal self becomes more congruent. Evidences of children's self-concepts and ideal selves were obtained by means of the Q-sort which was administered three times during the study, in December, February, and May, 1954–1955. Q-correlation coefficients of congruency of self-concept and ideal self were recorded for each child and transformed into $z$ scores. Mean $z$ scores and corresponding $r$ correlation coefficients were computed for each sex group, each classroom, and for each of the three times data were collected, and are reported in Table 3.

Analysis of variance tests of significance were made using the data in Table 3 to determine whether differences in congruency of children's self-concepts and ideal selves were significant for each of the four independent variables and for sex group and time. The results of these tests of significance are given in Table 4.

The results of the analysis of variance reported in Table 4 show that differences in congruency of children's self-concepts and ideal selves due to

TABLE 3. Mean $z$ Scores and $r$ Correlation Coefficients of Congruency of Children's Self-Concepts and Ideal Selves

| Teacher and Class | *N* | *Sex* | I Dec. $z$ | I Dec. $r$ | II Feb. $z$ | II Feb. $r$ | III May $z$ | III May $r$ | Mean $z$ and Corresponding $r$: $z$ | Mean $z$ and Corresponding $r$: $r$ |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 19 | Boys | .382 | .36 | .537 | .49 | .618 | .55 | .512 | .47 |
| A | 19 | Girls | .418 | .40 | .675 | .59 | .633 | .56 | .575 | .52 |
| | Mean $z$ and $r$ | | .400 | .38 | .606 | .54 | .625 | .55 | .544 | .50 |
| B | 10 | Boys | .451 | .42 | .485 | .45 | .503 | .46 | .480 | .45 |
| B | 5 | Girls | .691 | .60 | .835 | .68 | .829 | .68 | .785 | .67 |
| | Mean $z$ and $r$ | | .571 | .51 | .660 | .58 | .666 | .58 | .632 | .56 |
| C | 13 | Boys | .604 | .54 | .565 | .51 | .662 | .58 | .610 | .54 |
| C | 14 | Girls | .578 | .52 | .606 | .54 | .688 | .60 | .624 | .55 |
| | Mean $z$ and $r$ | | .591 | .53 | .585 | .53 | .675 | .59 | .617 | .55 |
| D | 23 | Boys | .526 | .48 | .625 | .55 | .688 | .60 | .613 | .55 |
| D | 18 | Girls | .641 | .56 | .815 | .67 | .827 | .68 | .761 | .64 |
| | Mean $z$ and $r$ | | .584 | .53 | .720 | .62 | .758 | .64 | .687 | .60 |
| J | 15 | Boys | .420 | .40 | .405 | .38 | .382 | .36 | .402 | .38 |
| J | 16 | Girls | .545 | .50 | .618 | .55 | .679 | .59 | .614 | .55 |
| | Mean $z$ and $r$ | | .482 | .45 | .512 | .47 | .530 | .48 | .508 | .47 |
| K | 13 | Boys | .360 | .35 | .355 | .34 | .303 | .29 | .339 | .33 |
| K | 18 | Girls | .554 | .50 | .493 | .46 | .498 | .46 | .515 | .47 |
| | Mean $z$ and $r$ | | .457 | .43 | .424 | .40 | .400 | .38 | .427 | .40 |
| L | 13 | Boys | .639 | .56 | .705 | .61 | .770 | .65 | .705 | .61 |
| L | 10 | Girls | .432 | .41 | .657 | .58 | .688 | .60 | .592 | .53 |
| | Mean $z$ and $r$ | | .536 | .49 | .681 | .59 | .729 | .62 | .648 | .57 |
| M–N | 12 | Boys | .428 | .40 | .399 | .38 | .582 | .52 | .470 | .44 |
| M–N | 21 | Girls | .475 | .44 | .481 | .45 | .666 | .58 | .541 | .49 |
| | Mean $z$ and $r$ | | .452 | .42 | .440 | .41 | .624 | .55 | .505 | .47 |

TABLE 4. F Ratios and Probabilities of Significant Differences for Sex, Time, and Four Independent Variables on Congruency of Children's Self-Concepts and Ideal Selves

| Basis of Comparison | *F* | *p* |
|---|---|---|
| Sex | 9.63 | .01 |
| Time | 3.71 | .05 |
| Climate | 2.12 | |
| Child Study | 8.65 | .01 |
| Teacher Acceptance | 2.71 | |
| Grade Level | 7.52 | .01 |

differences in sex are significant at the 1 per cent level. Differences in congruency at different points in time are significant at the 5 per cent level. These findings sustain hypotheses 1 and 2, for they show that girls generally register significantly greater self–ideal self congruency than do boys, but all children reveal significant changes toward greater congruency through time, December to May.

In Table 4 the report of the tests of significant differences relating to the four independent variables shows that differences in children's self–ideal self congruency are significant at the 1 per cent level for the variables of teacher participation in child study and grade level. Differences due to climate and teacher acceptance of self and others are not significant. Sixth grade children and those taught by teachers who have completed child study, then, reveal significantly greater self–ideal self congruency than do fourth grade children and those taught by non-child study teachers, respectively.

A further analysis was made to determine the influence of the independent variables on the proportionate number of increases and decreases in children's self-ideal self congruency. This was accomplished by tabulating for each child in each classroom the frequencies of increase and decrease in self–ideal self congruency as shown by a comparison of $z$ scores for three periods of time, December–February, February–May, and December–May. Chi square tests of independence revealed that the proportionate numbers of increases and decreases in children's self–ideal self congruencies produced results similar to those obtained by the analysis of variance treatments in Table 4. The chi square results were less clear cut, however, because the effects of magnitude of change were not included. These findings, then, support the general hypothesis only with respect to sixth grades and classrooms whose teachers have completed child study in which children reveal significantly greater change in self–ideal self congruency.

This research also sought to ascertain the relationships that exist between changes in children's self–ideal self congruencies and changes in children's school achievement as shown by differences in scores made on the California Achievement Tests which were administered to the four sixth grade classrooms before and after the period of this study. Evidences of changes in children's self–ideal self congruency were obtained by computing for each child the $z$ score difference between his $z$ score self–ideal self congruency in December and his $z$ score self–ideal self congruency in May. Mean product-moment correlation coefficients for the four sixth grade classrooms for relationships between changes in children's self-ideal self congruency and changes in academic achievement were .09 for language, —.06 for reading, .16 for arithmetic, and .08 for total achievement. Since these coefficients do not differ significantly from zero, it is concluded that there are no significant relationships between changes in children's self–ideal self congruencies and changes in their school achievement.

Finally, this research sought to determine the relationship that exists between changes in children's self-ideal self congruency and changes in their acceptance by the peer group. Measures of change in self–ideal self congruency used in studying this relationship were the same differences in $z$ score congruency, December and May, which were used previously in correlating this factor with school achievement. Measures of change in peer group acceptance were obtained from the answers children made to a twenty-sixth question on the sociometric questionnaire: Who are your very best friends in this class? Change in peer acceptance was defined as the difference between the number of times a child was named as a friend in December and the number of times he was named in May. Mean product-moment correlations for the relationships of changes in children's self–ideal self congruency and changes in their peer acceptance were .04 for fourth grades and .03 for sixth grades. Mean correlations for the relationship of $z$ score self–ideal self congruency and peer group acceptance at the end of the study in May, 1955, were .17 for fourth grades and .27 for sixth grades. These findings show that there is a lack of relationship between changes in children's self–ideal self congruency and changes in their acceptance by peers and that children with greater self–ideal self congruency are not chosen more frequently by their peers. Thus, the lack of relationships between change in children's self-ideal self congruency and changes in their school achievement and peer group acceptance indicates a lack of support in this study for hypothesis 3.

## Discussion

In this research it was posited that increase in self–ideal self congruency facilitates self-development and is evidence of the improved self-adjustment of the individual. Since a major purpose of education is to promote individual self-development, the challenge of helping individuals increase their self–ideal self congruency would seem to be of crucial importance to education.

It is reassuring to note, therefore, that according to the findings of this study children in general do achieve increased self–ideal self congruency during a period of time that they are in attendance at school. Two explanations for this increase are suggested. First, it may be influenced by the general pattern of children's growth and development. It seems reasonable to postulate that as children mature, their self-concepts become more stable and consistent and their self-concepts and ideal selves increase in congruency. This would not deny, of course, the importance of specific factors in the environment or life experience of the individual which would be likely to influence his self–ideal self congruency irrespective of the matura-

3. Sixth grade children and children whose teachers have completed child study show significantly greater self–ideal self congruency than do children, respectively, in fourth grade and those whose teachers have never participated in this child study program.
4. There is little or no relationship between changes in children's self–ideal self congruency and (a) changes in their school achievement, and (b) changes in their acceptance by peers.

In the discussion of the findings it was hypothesized that changes in self–ideal self-congruency of children in general may be a manifestation of growth and development and may reflect the positive influence of school experience. The greater self–ideal self congruencies of girls contrasted with boys may relate to their greater physical maturity and may be influenced by the school's feminine mores and codes which tend to favor girls. The greater self–ideal self congruencies of children whose teachers have participated in child study is evidence that this program benefits not only teachers but children as well.

## REFERENCES

1. Brandt, R. M. Accuracy of self-estimate: a measure of self-concept reality. *Genet. Psychol. Monogr.*, in press.
2. Chodorkoff, B. Adjustment and discrepancy between the perceived and ideal self. *J. clin. Psychol.*, 1954, **10**, 266–268.
3. Jersild, A. T. *In search of self*. New York: Teachers College, Columbia Univer., 1952.
4. Lepine, L. T., & Chodoroff, B. Goal setting behavior, expressed feelings of adequacy and the correspondence between perceived and ideal self. *J. clin. Psychol.*, 1955, **11**, 395–397.
5. Perkins, H. V. Teachers' and peers' perceptions of children's self-concepts. *Child Develpm.*, 1958, **29**, 203–220.
6. Reed, H. J. An investigation of the relationship between teacher effectiveness and the teacher's attitude of acceptance. *J. exp. Educ.*, 1953, **21**, 277–325.
7. Withall, J. G. The development of a technique for the measurement of social-emotional climate in the classroom. *J. exp. Educ.*, 1949, **17**, 347–361.

# 30

# A System of the Psychomotor Abilities

## — *J. P. Guilford*

*In this article the author attempts to develop a system of psychomotor abilities by seeking common characteristics among different body responses and movements. The result is a two-way classification of psychomotor abilities, in terms of the type of ability and of the part of the body involved. Also presented are some of the characteristic tests used to measure these abilities. The system is intended by the author to be viewed as a theory of psychomotor abilities.*

DURING THE past three years, the author has had occasion to make a systematic and comprehensive survey of the so-called primary mental abilities that have been found by factor analysis in the general area of intelligence. The result has been called a "structure of intellect," in which the known intellectual factors have found places and a number of unknown factors have been predicted by the system.[1] The system is in general an hierarchical one, with major and minor groups of abilities. Some interesting principles of classification of the factors suggest new conceptions of mental functioning as well as interrelationships of abilities within a functioning individual. New concepts of thinking that were suggested promise to be use-

Reprinted from *American Journal of Psychology*, 1958, **71**, 164–174, by permission of the publisher.

[1] J. P. Guilford, The structures of intellect, *Psychol. Bull.*, **53**, 1956, 267–293; A revised structure of intellect, *Rep. psychol. Lab.* No. 19, University of Southern California, 1957.

ful in accounting for such complex behavior as problem-solving, creative thinking, strategy, judgment, and decision-making.

The apparent significance and usefulness of such an organization of the factors of intellect suggested that it might be profitable to attempt a similar survey of the psychomotor abilities, concerning which we have learned a great deal during the past few years, largely through the efforts of Fleishman and Hempel.[2] Proceeding as in the previous study, the author first typed each factor reported in the literature on a separate card, together with its apparent properties and names of some tests that tend to characterize it. With the high degree of maneuverability provided by this procedure, it was possible to arrive at classes of factors and to find that a kind of two-way classification seems called for. After a number of revisions, the end-product appears as in Table 1, where it will be seen that the known psychomotor factors can be placed in a single plane in the form of a matrix. The general nature of the matrix will be explained, the factors will be defined, and their most characteristic tests will be listed. Some apparent implications of the system will be mentioned.

TABLE 1. Matrix of the Psychomotor Factors, with Columns for Kinds of Abilities and Rows for Parts of the Body Involved.

| Part of Body Involved | Type of Ability | | | | | | |
|---|---|---|---|---|---|---|---|
| | Strength | Impulsion | Speed | Static Precision | Dynamic Precision | Coördination | Flexibility |
| Gross | General strength | General reaction-time | | Static balance | Dynamic balance | Gross bodily coördination | |
| Trunk | Trunk-strength | | | | | | Trunk-flexibility |
| Limbs | Limb-strength | Limb-thrust | Arm-speed | Arm-steadiness | Arm-aiming | | Leg-flexibility |
| Hand | | Tapping | | | Hand-aiming | Hand-dexterity | |
| Finger | | | Finger-speed | | | Finger-dexterity | |

## Concerning Analyses of Psychomotor Abilities

Before the system of psychomotor factors is described, some pertinent circumstances should be mentioned regarding the factor-analysis studies that have been conducted in this area. There are some special problems involved, and some studies have shortcomings at some points for the purpose of investigating basic abilities.

---

[2] E. A. Fleishman, Dimensional analysis of psychomotor abilities, *J. exp. Psychol.*, **48**, 1954, 437–454; E. A. Fleishman & W. E. Hempel, Jr., A factor analysis of dexterity tests, *Personnel Psychol.*, **7**, 1954, 14–32; Factorial analysis of complex psychomotor performance and related skills, *J. appl. Psychol.*, **40**, 1956, 96–104; Hempel & Fleishman, A factor analysis of physical proficiency and manipulative skills, *J. appl. Psychol.*, **39**, 1955, 12–16.

Generally, three kinds of tests have been used: tests of physical fitness, apparatus-tests, and printed tests. Most commonly, tests of only one kind have been used in each analysis. Tests of physical fitness involve the whole body, or large portions of it. They are given without much in the way of equipment, and they are most often given in physical-education and athletic settings. Printed tests are adapted to group-testing, the kinds of movement involved being fairly well restricted to finger and hand. Fortunately, apparatus-tests can involve movements all the way from finger-reactions to gross, bodily activity, thus overlapping both the other two types. They have sometimes been administered along with the other two types for analytical purposes, but not as often as would have been desirable. Some factors have been found in common to apparatus-tests and the other two kinds, but usually only when the same muscle-groups and similar movements are involved.

One of the circumstances that has operated against more rapid progress in differentiating the factors is that it is sometimes difficult to design tests that involve isolated bodily members, for example, tests involving finger-movements separate from hand- and wrist-movements, and these separate from arm-movements. Even to the extent that such separation of functions would be possible, efforts have not often been made to effect it by means of experimental controls.

Another somewhat bothersome circumstance is that, in many analyses, two scores from each of one or more tests have been included. An example is a measure of the performance of the left hand in addition to a measure of the performance of the right hand. Minor variations of certain tests provide another instance. The result has been that such tests achieve over-inflated loadings as compared with other tests, due to the confounding of specific variances with common-factor variances.

In spite of these difficulties and qualifications, it is possible to see that certain factors are found with some degree of invariance from one analysis to another. Where two factors are loaded on one or more of the same tests, the hypothesis regarding identity has empirical support. Where one or more similar, but not identical, tests are in common, the support is somewhat weaker. Some factors have been reported in only one analysis, but, where such factors seem to fit into the system, they are mentioned not only as potential contenders for membership in the family, but as holding certain unique places within the family.

The entire system should be regarded as a kind of theory of psychomotor abilities, with each cell as an hypothesis. Some cell-hypotheses have con-

There are enough differences in reaction-time for different muscle-groups, however, to suggest that there may also be regional reaction-time abilities.[8] The sense-modality of the stimulus seems to make no difference as far as present information goes.

*Limb-thrust*[9]
Vertical jump
Standing broad-jump
Shot-put (variable weights)
Short dash
Bar-vault

This may be a confounding of two factors, one for arms and one for legs. If there are two factors, there is probably much correlation between them. The explosive aspect of movements, which has suggested the term *impulsion* for this class of abilities, is most evident in this group of tests.

*Tapping*[10]
Tapping, index finger
Tapping, turnbuckle (*S* turns turnbuckle with fingers.)
Tapping, four fingers in turn
Pegs, handicap (*S* removes and replaces pegs in holes.)

It is not certain that the tapping factor, which now seems confined very much to finger-movements, belongs in the same column with the other impulsion-factors, since it also involves *repetition* of initiated movements. Presumably the recovery-time between impulsions should have much bearing upon the scores. Until tapping tests are applied to different bodily parts, and more extensively analyzed with impulsion-tests, we cannot answer this question. The present information is that tapping scores from the hand do not correlate with scores for reaction-time also obtained from the hand.[11]

*Articulation-speed*[12]
Maximal rate of oral reading
Normal rate of oral reading
Speed of articulation (repeating a consonant)
Spelling (in a dictation test)

Tests related to this factor are formally like tapping tests, particularly the third in the list. The articulation-factor thus seems to fall under *impulsion,* or it might go with a new tapping category if such a category is found to be needed.

---

[8] R. H. Seashore, R. Starmann, W. E. Kendall, & J. S. Helmick, Group factors in simple and discrimination reaction time, *J. exp. Psychol.*, **29**, 1941, 346–349.

[9] A. J. Wendler, A critical analysis of test elements used in physical education, *Res. Quart. Amer. Ass. Hlth phys. Educ.*, **9**, 1938, 64–76; McCloy, *op. cit.*, 173–182; Carpenter, *op. cit.*, 34–40; Shapiro, *op. cit.*; Hempel & Fleishman, *op. cit.*, 437–454.

[10] Wittenborn, *op. cit.*, 395–409.

[11] Seashore, Starmann, Kendall, & Helmick, *op. cit.*, 346–349.

[12] J. B. Carroll, A factor analysis of verbal abilities, *Psychometrika*, **6**, 1941, 279–307.

### *Motor-speed factors*

The hypothesis that there is a general motor-speed factor has been taken seriously by a number of investigators, but such an ability has not been clearly demonstrated.[13] Rimoldi reports evidence that might be interpreted as supporting such an hypothesis, but his tests were designed to measure a trait of personal "tempo" instead of speed.[14] The instructions were for S to respond with his natural speed, but he evidently could know that his reactions were being timed. Harrison found an average correlation of 0.44 between natural and maximal rates of performance on various tests, hence Rimoldi's factors might be speed-factors.[15]

- *Arm-speed*[16]
  - Ten-target aiming (*S* aims at ten targets in succession.)
  - Two-plate tapping (*S* taps alternately on two plates some distance apart.)
  - Rotary aiming (*S* touches succession of buttons with finger tip.)
- *Hand-speed*[17]
  - Pin-stick (*S* loops string around succession of pegs.)
  - Tapping, large (*S* writes dots in scattered circles, each 7/16 in. in diameter.)
  - Tapping, small (*S* writes dots in circles ⅛ in. in diameter.)
  - Santa Ana finger-dexterity (*S* removes each square peg from hole, turns it 180°, and resets it.)
- *Finger-speed*[18]
  - Log-book accuracy (*S* marks spaces in an answer-sheet.)
  - Tapping, small
  - Square-marking (*S* makes a cross in each square.)
  - Marking-accuracy (another test that involves marking an answer sheet.)

### *Static precision-factors*

- *Static-balance*[19]
  - Bending (*S* stands on ball of foot, eyes open.)
  - Foot-balance II (*S* stands with right foot held behind left knee.)
  - Two-foot rail-balance (*S* walks, heel-to-toe, on 1-in. board.)

---

[13] R. Harrison, Personal tempo and the interrelationships of voluntary and maximal rates of movements, *J. gen. Psychol.*, **24**, 1941, 343–379.

[14] Rimoldi, *op. cit.*, 283–303.

[15] Harrison, *op. cit.*, 343–379.

[16] Fleishman, *op. cit.*, 437–454.

[17] Rimoldi, *op. cit.*, 283–303; Fleishman & Hempel, *op. cit.*, 14–32.

[18] J. P. Guilford & J. I. Lacey (Eds.), *Printed Classification Tests*, AAF, *Aviation Psychology Research Program Reports*, No. 5, 1947, 830–831; Fleishman, *op. cit.*, 437–454; Fleishman & Hempel, *op. cit.*, 14–32; J. P. Guilford & P. R. Christensen, A factor analytic study of verbal fluency, *Rep. psychol. Lab.*, No. 17, University of Southern California, 1956.

[19] R. I. Bass, An analysis of the components of tests of semicircular canal function and of static and dynamic balance, *Res. Quart. Amer. Ass. Hlth phys. Educ.*, **10**, 1939, 33–52; Hempel & Fleishman, *op. cit.*, 12–16.

*Arm-steadiness*[20]

- Track-tracing (*S* moves stylus in slot at arm's length.)
- Steadiness-aiming (*S* keeps delicately balanced stylus in hole at arm's length.)
- Cox eye-board (*S* threads cord through eyes in pegs in row.)
- Hex-nut steadiness (*S* stacks series of small, hexagonal nuts.)

Although some of the tests in the lists for factors in this category involve some movement, for the most part they have to do with holding bodily members in position.

*Dynamic precision-factors*

*Dynamic balance*[21]

- Jump and click heels together (while in the air)
- Jump and balance (*S* jumps upon edge of box and maintains his balance.)
- Rate of jump (*S* jumps up stairs one step at a time, feet together, then jumps backwards down stairs.)

*Arm-aiming*[22]

- Pursuit-aiming II (*S* traces through series of small openings in a maze-pattern.)
- Pursuit-aiming I (*S* traces small circles, putting a dot in each circle.)
- Punch-board (*S* puts pins through tiny holes rapidly.)

*Hand-aiming*[23]

- Irregular dotting pursuit (*S* places dots in small circles.)
- Circle-dotting (*S* puts three dots in each small circle.)
- Rotary aiming (*S* touches each button in a circle.)

*Coördination factors*

*Gross bodily coördination*[24]

- Cable-jump (*S* jumps through a cable that he holds.)
- Rotary pursuit
- Hurdle-jump
- Rudder-control (*S* corrects balance of his seat by manipulating controls with his feet.)
- Complex coördination (*S* operates stick and rudder-controls to effect patterns of lights to match others.)

This factor apparently involves coördinated activity of muscles of arms and legs, and perhaps also the trunk. Whether we shall find separate arm- and leg-factors remains to be seen. From one study it has been suggested that there are two coördination-factors involving gross bodily parts.[25] The one mentioned above is considered a rough-adjustment ability and the other a fine-adjustment ability. The writer is inclined to regard the latter as a dynamic-precision factor, which would belong in the previous category.

---

[20] Hempel & Fleishman, *op. cit.*, 12–16.
[21] Hempel & Fleishman, *op. cit.*, 12–16; Fleishman, *op. cit.*, 437–454.
[22] Fleishman, *op. cit.*, 437–454; Hempel & Fleishman, *op. cit.*, 12–16.
[23] Hempel & Fleishman, *op. cit.*, 12–16.
[24] Guilford & Lacey, *op. cit.*, 830–831; Shapiro, *op. cit.*; Fleishman, *op. cit.*, 437–454.
[25] Fleishman & Hempel, *op. cit.*, 96–104.

*Hand-dexterity*[26]
  Marble-board (*S* places marbles in a groove.)
  Santa Ana finger-dexterity (described above)
  *VDL* rings (*S* removes small rings from small pole.)
  Dowel-manipulation (*S* arranges suspended pieces to make a continuous dowel.)

*Finger-dexterity*[27]
  Purdue pegboard, both hands
  Purdue pegboard, assembly
  O'Connor finger-dexterity
  Nut and bolt (*S* completes nut-and-bolt assemblies.)

*Flexibility Factors*

Those who are concerned with physical education have recognized for a long time that flexibility of movement is important in athletic training. It depends very much upon the looseness of the joints. An important problem for the factorist is to determine whether flexibility is a general property, about equally apparent in all joints of the individual, or whether there is room for localized flexibility depending upon the bodily parts. Two such localized abilities have been found.

*Trunk-flexibility*[27]
  Back down the wall (*S* bends over backwards, hands on wall.)
  Table-vault (Placing hands on table, *S* vaults over it landing on his feet.)

In both tests just mentioned, it would seem that success depends upon the freedom of the spinal column to bend and twist.

*Leg flexibility*[28]
  Toe-touching (*S* touches toes with fingers without bending knees.)
  Kicking-height
  Leg-bend (*S* squats slowly on right leg.)

## Discussion

### The Nature of the Whole-body Factors

Having the gross factors mentioned and described in part, we should now examine them to see what kinds of ability they seem to represent. Two of them—*general strength* and *general reaction-time*—genuinely appear to be general factors that can be attributed to positive intercorrelations of tests applied to various bodily parts. Other factors in the first row of Table 1 seem rather to represent properties of total-body activities—*static balance, dynamic balance,* and *gross bodily coördination.* That is, the ability in each case pertains to a total-body performance of some kind, or to simultaneous involvement of two or more regions. They do not seem to owe their existence to the correlations of tests across regions, though that may be found to be the situation when more information is available. No general speed-factor has been demonstrated, although there is some evidence for it. No general flexibility-factor has been found, probably because not enough kinds of flexibility-tests have been explored. If a total-body flexibility-factor exists, it would seem to be of the general type.

---

[26] Wittenborn, *op. cit.*, 395–409; C. H. Goodman, The McQuarrie test for mechanical ability: II. Factor analysis, *J. appl. Psychol.*, **31**, 1947, 150–154; Fleishman, *op. cit.*, 437–454; Fleishman & Hempel, *op. cit.*, 14–32; Hempel & Fleishman, *op. cit.*, 12–16.

[27] Fleishman, *op. cit.*, 437–454; Fleishman & Hempel, *op. cit.*, 14–32; Hempel and Fleishman, *op. cit.*, 12–16.

[28] Hempel & Fleishman, *op. cit.*, 12–16.

The existence of regional psychomotor-factors in addition to general ones calls for some speculation. One reasonable hypothesis might be that the general factors are relatively more dependent on heredity and the regional ones relatively more on experience. We should expect the pianist to develop strong fingers; the cyclist, strong legs; and the coal-fireman strong shoulders, arms, and trunk. Since, however, the anatomical properties on which psychomotor abilities undoubtedly depend may vary from one region to another due to heredity, we properly should not stress too much a nature-nurture difference between general and regional factors.

### Other Possible Psychomotor Factors

Undiscovered primary psychomotor abilities are suggested by the vacant cells in Table 1. It may be surmised that the chief reason for the vacancies is the lack of interest in such abilities, for example, those that pertain to the trunk. Vocational-fitness problems have determined some selection of abilities that have been studied. Instructors of physical education and athletic coaches have been interested in flexibility-factors whereas psychologists have not. The psychologist's interest should be sufficiently broad and impartial to direct his study systematically to the whole field of psychomotor abilities. Not indicated by the table is the possibility of a whole array of factors pertaining to vocal movements, which should also be of interest to investigators in the field of speech.

From investigators in the field of physical education, we have suggestions of a few factors that do not find ready places in Table 1. McCloy and Young mention a factor of *muscular endurance* and a factor of *circulatory-respiratory endurance.*[29] It is possible that these traits are composites or syndromes of physiological characteristics, combined with a motivational factor of endurance, which is the tendency to continue to work toward a goal in spite of discomfort or pain. The same authors speak of a factor of *agility*, which is shown in such activities as open-field running and dodging. It is possible that this is a syndrome of certain impulsion-factors combined with factors of coördination and speed. The concept of *power* also seems to be a syndrome of speed or impulsion and strength, for it has to do with forceful motion. These are all in the nature of hypotheses to be tested by a combination of experimental and factor-analysis methods.

---

[29] C. H. McCloy & N. D. Young, *Tests and measurements in health and physical education*, 3rd ed., New York, Appleton-Century-Crofts, 1954.

## Some General Implications of the System of Psychomotor Factors

A number of interesting questions are raised by reflecting upon the psychomotor factors, and upon the principles that seem to apply to them, in relation to anatomical and physiological features of the human organism. The obvious dependence of such abilities upon anatomical and physiological traits calls for study of the intercorrelations between these modalities of traits. Presumably this kind of study has been done by those interested in development of athletic skills, but probably not systematically and in terms of factors. The properties of bones and muscles, and the manner in which they are put together, should have much explanatory significance in accounting for many psychomotor abilities.

Some of the distinctions between classes of psychomotor abilities should be of interest to the physiological psychologist. Neurological parts and properties might account for many of the distinctions that appear in Table 1. For example, what brain-structures should be most credited for impulsion *versus* speed, for precision, for coördination? Why should there be a difference between precision of muscular control in the static condition and precision of muscular control in motion? What are the roles of visual *versus* static-sense control in the factors of static and dynamic balance? These are a few of the unanswered questions concerning motor control, all of which possibly have received attention but which seem to be highlighted by the way in which the results of factor-analysis are organized in Table 1.

# 31

# The Learning of Motor Skills as Influenced by Knowledge of Mechanical Principles

## — *Frances M. Colville*

*Does teaching the principle underlying a motor skill facilitate immediate learning of the skill? And does learning the principle positively affect acquiring a different or more complicated skill, if the same principle is applicable? In answer to these questions, the author reports an investigation, employing ingeniously contrived tasks, that assesses the effects of knowledge of mechanical principles on task performance.*

MANY GENERAL principles of mechanics are relevant to the teaching, learning, and performance of activities included in the physical education curriculum. These principles may describe the motion of objects such as balls and racquets, movement of the body itself, or a combination of both.

The question has been raised as to whether knowledge of these general principles and an understanding of their application to these activities will facilitate the learning and improve the performance of pupils in physical education classes. Also of interest are questions concerning the influence of knowledge of these principles as applied to one activity upon subsequent learning of other activities to which the same principles apply. Results of a few experiments suggest that a pupil who understands a principle related

Reprinted from *Journal of Educational Psychology*, 1957, 48, 321–327, by permission of the author and the publisher.

to one skill may master a related skill more readily than the pupil whose experience has been restricted to specific instruction in technique without explanation of pertinent principles.

With reference to the teaching of a motor skill in which a specific principle of mechanics is involved, then, two questions have been raised:

1. What is the effect of knowledge of a principle upon immediate learning of a skill to which the principle applies?
2. What is the effect of knowledge of a principle learned in relation to one skill upon subsequent learning of a different or more complicated skill to which the same principle is applicable?

Reported investigations bearing on the acquisition and transfer of principles are limited, and although most authorities agree that general principles are probably transferable, little strong experimental evidence is available to support such a claim. Apparently, little or no experimentation has been done in the field of physical education on the effect of knowledge of principles of mechanics in learning situations involving body movement and large muscle activities.

Judd (*7*), and Hendrickson and Schroeder (*4*) reported that knowledge of the principle of refraction was beneficial in dart throwing and rifle shooting at a target under water. Cox (*2*) found that knowledge of principles involved in manipulating electrical equipment facilitated the assembling and stripping of such equipment. Other experimenters have reported similar results from investigations involving mathematics, mental games, spelling, card tricks, and mechanical puzzles (*13, 11, 15, 10, 14, 8, 2*). In contrast to these findings are a number of investigations in which the method of including instruction in the principles involved did not appear to be more effective. In studying the effect of drill versus the learning of generalizations related to addition and subtraction, Olander (*9*) found that the two methods appeared to be equally effective. Babitz and Keys (*1*) found similar results with college chemistry classes, and Hendrix (*6*) reported only one difference, significant at the 12% level, in her experiment with teaching algebraic generalizations.

With the exception of the experiments by Judd, Hendrickson and Schroeder, and Cox, all of the reported investigations have dealt with various types of learning other than motor learning, and only Judd's experiment involved skills requiring the coordinated use of large muscles of the body as well as fine coordinations of the fingers. Judd, and Hendrickson and Schroeder, while reporting a difference in favor of instruction in principles, included this instruction in such a way that additional time was allotted to the group receiving the instruction. Thus, the experimental groups had the advantage of additional time as well as additional information. Furthermore, Judd supplied no statistical analysis of his early experiment, and the other differences reported were significant at a statistical level below that which is usually considered acceptable.

In the present investigation, an attempt was made to control these ambiguities by using an experimental design which insured that both the experimental and the control groups spent equal amounts of time in learning and performing and by considering only those differences which were significant at the 5% level of confidence or above.

## Method

### Purpose of the Study

It was the purpose of this study to investigate certain questions related to the teaching of physical education activities in which specific principles of mechanics are involved. This problem was approached by: (*a*) selecting three principles of mechanics which are pertinent to motor skills; (*b*) selecting three motor skills, each of which utilizes one of the principles; (*c*) establishing for each skill two comparable groups of Ss, one of which was taught without reference to the principle involved, and the other of which was taught to understand and apply the principle; (*d*) comparing for each skill the performance of the two groups, one of which spent the entire time practicing the skill without reference to the principle involved, and the other of which spent part of the time practicing the skill and part of the time in learning the principle; (*e*) comparing for each skill the performance of the two groups in a similar or more complicated form of the skill to which the same general principle applies.

### Selected Principles and Skills

The selected principles and the skills in which they are utilized were as follows:

*Principle I*: The angle of incidence is approximately equal to the angle of reflection. Skill: Rolling a ball against a surface, or surfaces, from which it would rebound.

*Principle II:* In stopping a moving object, the force opposing the momentum must be equal to the force of the momentum, and if the object is to be caught, this momentum must be dissipated by reducing the resistance of the catching surface. Skill: Catching a tennis ball in a lacrosse stick and catching a badminton bird on a tennis racquet.

*Principle III:* An object set in forward motion through the air by an external force is acted upon by the force of the momentum and by gravital acceleration. Skill: Archery.

**Procedure**

Three parallel experiments were devised, each of which was designed to investigate performance of one of the three skills. The Ss for all three experiments were undergraduate women students at the University of Southern California. There were 36 Ss in the Ball Rolling experiment, 40 in the Catching experiment, and 42 in the Archery experiment. (The Ss in the Archery experiment included some men whose scores were treated separately from those of the women.) For each of the experiments the Ss were divided into two groups, one of which spent the entire time learning and practicing the skill, and the other of which spent part of the time learning about the principle and the rest of the time learning and practicing the skill. Each experiment consisted of two tests. During the First test of the Ss of the nonprinciple group learned the skill and practiced. The Ss of the principle group learned the principle, and the skill, and practiced. The total amount of time spent by both groups was the same. During the Second test both groups practiced a similar or more complicated skill. Each group spent the same amount of time and no mention was made of the principle involved.

*The Ball Rolling experiment.* The Ball Rolling experiment was designed to illustrate the principle describing the rebound of a ball. The apparatus consisted of a modified pinball machine plunger, a squash ball, a felt-covered table surface four feet wide and six feet long, and a target made of colored construction paper. The overall size of the target and each of its scoring divisions was determined by recording the hits made by a group of comparable Ss who were not used in the experiment. These Ss rolled the ball so that it would rebound from one side of the table as close as possible to the center of the adjacent side which was marked with a square of white paper. The rest of the side was covered with brown wrapping paper, and spots on this paper to which the balls rolled were marked in red pencil. This scoring sheet provided the basis for establishing a six-standard deviation scale, and from this scale the scoring areas of the target were derived.

The Ss were asked during the First test to roll the ball against one side of the table in such a manner that it would rebound to the target placed against the adjacent side. For the Second test they were asked to roll the ball so that it would rebound from two adjacent sides to the target set at right angles on the third side.

*The Catching experiment.* The catching experiment consisted of two tests each involving the skill of catching a moving object. The First test involved catching a tennis ball in a lacrosse stick. For the Second test, they caught a badminton bird on a tennis racquet. The apparatus for this experiment, in addition to balls, birds, crosse, and racquet, included a homemade device for projecting the tennis ball and the badminton bird. This device proved to have high reliability.

*The Archery experiment.* The archery experiment was planned to illustrate the principle of gravital acceleration and its action on arrow flight and the relation of line of flight to line of sight. The First test involved shooting from 20 yards. The Second test included ends shot from 30 yards, from 40 yards, and a Junior Columbia Round.

*Analysis of data.* In each experiment, the reliability of the scores was established by use of the split-half technique, the odd scores being correlated with the even scores. Use of a Spearman-Brown Prophecy Formula for predicting reliability of the whole test resulted in coefficients which ranged from .81 to .94 and thus it was felt that these scores were sufficiently reliable to permit further analysis and comparison.

All of the data were analyzed by means of analysis of variance using Edwards' (3) technique for repeated measurements of the same subjects. This technique tests differences between methods, between trials, and between the interaction of trials and methods. In each experiment the initial ability of the two groups, as evidenced by preliminary trials, was tested and found not to differ significantly.

## Findings

In the Ball Rolling experiment, no significant difference in the performance levels of the two groups was found on either test. A difference between trials significant at the 1% level of confidence was found on both tests.

In the Catching experiment, no significant differences were found in performance levels, and a difference between trials significant at the 5% level of confidence was found only on the First test.

In the Archery experiment no significant difference in performance levels was found. A difference significant at the 1% level for men and at the 5% level for girls was found on the Second test at all distances except for forty yards.

The following general findings were noted:

1. A significant amount of learning took place under both methods of instruction.
2. This learning was not only similar in amount but also in pattern. Exceptions occurred in the Second test of the Catching experiment where no significant increase in scores occurred in either group, and in the Junior Columbia Round for men in the Archery experiment, where both groups increased in scoring ability, but the group which had been taught the principle was significantly better at twenty yards.

These findings are in general agreement with those of some published investigations and are opposed to those of some others. However, most of the reported experiments have dealt with various types of learning other than motor learning. The ability to use understood principles in performing a motor skill presents a complex problem since more than one type of learning is involved. As Ragsdale (*12*) has pointed out, a learner may understand a principle when applied to inanimate objects, but he may not understand and be able to apply it to his own movements. In general, it seems that the findings of this study support Ragsdale's observation, at least in the initial stages of acquiring an unfamiliar motor skill.

## Conclusions

Within the limitations of the three parallel experiments which constitute the present investigation, there is no evidence:

1. That instruction concerning mechanical principles utilized in the performance of a motor skill facilitates the initial learning of the skill to any greater extent than an equivalent amount of time spent in practicing the skill.
2. That such knowledge facilitates subsequent learning as evidenced in the performance of a similar or more complicated skill to which the same principle is applicable.

However, since it appears that some part of the learning period may be devoted to instruction concerning general principles without detriment to the motor learning of the students, it would seem desirable to include such instruction in order to provide this additional opportunity for acquiring some related knowledge about principles of mechanics and the application of forces.

## REFERENCES

1. Babitz, M., & Keys, N. Experiment in teaching pupils to apply scientific principles. *Sci. Educ.*, 1939, 23, 367–370.
2. Cox, J. W. Some experiments on formal training in the acquisition of skill. *Brit. J. Psychol.*, 1933, 24, 67–87.
3. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
4. Hendrickson, G., & Schroeder, W. H. Transfer of training in learning to hit a submerged target. *J. educ. Psychol.*, 1941, 32, 205–213.
5. Hendrickson, G., & Schroeder, W. H. Transfer of training. In R. G. Kuhlen & G. G. Thompson (Eds.), *Psychological studies of human development.* New York: Appleton-Century-Crofts, 1952.
6. Hendrix, Gertrude. A new clue to transfer of training. *Elem. Sch. J.*, 1947, 48, 197–208.
7. Judd, C. H. Special training and general intelligence. *Educ. Rev.*, 1908, 36, 38–42.

RESEARCH ON motivation, retention, and transfer, as these relate to learning, is proceeding at a rapid pace. Motivation is an integral component of purposeful learning and is closely related to the efficiency with which students learn. Teachers are properly concerned also with retention (remembering what was learned) and transfer (the ability to use what was learned and is remembered in meeting new or different situations with greater effectiveness). Part VI contains six selections illustrating research in these three areas.

One of the advantages cited by advocates of learning by self-discovery is that it generates intrinsic interest in inquiry (see, for instance, reading number 19). In the lead-off article in this part Bert Kersh reports an experiment designed to ascertain this motivational effect. Kersh is associate director of Teaching Research in the Oregon State System of Higher Education. On the basis of his findings he concludes that learning by self-discovery is superior to learning with external direction only insofar as the former increases student motivation to pursue the learning task.

Marvin Shaw, formerly with the Massachusetts Institute of Technology, and now professor in the Department of Psychology at the University of Florida, reports in reading number 33 an investigation of motivational factors in cooperation and competition. The complex nature of motivational effects in cooperative and competitive task situations is evident from the results of his research.

In the next selection, David Ausubel reports an experiment based on a subsumption theory of learning. Ausubel, professor of education at the University of Illinois, employs "advance organizers" to provide relevant subsuming concepts in introducing meaningful but unfamiliar verbal material. Results of the investigation support his theoretical premises and have implications both for teaching procedures and for textbook organization.

Herbert Klausmeier, of the University of Wisconsin, and John Check, of Flint College (University of Michigan), report in reading number 35 a study of retention and transfer in a classroom setting. Using learning material that is meaningful and consistent with the learning abilities of the subjects, they found no significant differences in retention and transfer abilities among fifth-graders of low, average, and high intelligence. Interesting implications for encouraging efficient acquisition and subsequent retention and transfer of learning emerge from the study.

*In reading number 36, G. M. Haslerud and Shirley Meyers, both of the University of New Hampshire, examine the transfer effects of learning by self-discovery. Together with the logical arguments offered by Bruner supporting self-discovery (see reading number 19) and the research reported by Kersh (reading number 32), the findings of this study build a strong case for the self-discovery method of learning.*

*The last selection, "Transfer Effects of Learning a Verbal Generalization," is by Susan Ervin, who is associated with the Center for Human Learning at the University of California, Berkeley. Miss Ervin reports an investigation indicating that there are several conditions necessary before learning a verbal generalization will facilitate transfer.*

# 32

# The Motivating Effect of Learning by Directed Discovery — *Bert Y. Kersh*

*The effectiveness of learning by self-discovery as compared to the effectiveness of directed learning is a question of concern to all educators. The author of the research reported here addresses himself to this pertinent question. The results of his research support arguments for self-discovery on the basis of its powerful motivating effect. It is concluded that this method of learning motivates students to practice more, and thus to remember and transfer more, than they might if taught directly. The author's provocative discussion of his findings has important implications both for additional research and for classroom teaching.*

ADVOCATES OF the process of learning by directed discovery claim a number of advantages, most of which are included in a recent article by Bruner (1961). He has suggested that learning by discovery benefits the learner in four ways: it (*a*) increases the learner's ability to learn related material, (*b*) fosters an interest in the activity itself rather than in the rewards which may follow from the learning, (*c*) develops ability to approach problems in a way that will more likely lead to a solution, and (*d*) tends to make the material that is learned easier to retrieve or reconstruct.

Research evidence does not entirely support Bruner's arguments. One of the more recent reviewers, Ausubel (1961, p. 47), concludes "that most of the reasonably well-controlled studies report negative findings." However,

Reprinted from *Journal of Educational Psychology*, 1962, **53**, 65–71, by permission of the author and the American Psychological Association.

as is true in other areas of research, the evidence is somewhat equivocal, partly because it is difficult to equate studies in terms of the amount and kind of direction that is provided. The experimental subjects rarely if ever are required to learn completely without help, and the kind of help provided commonly differs. Consequently, there are studies which appear to be somewhat contradictory, such as Craig's (1956), in which the "directed" group learned and retained significantly more principles than the "independent" group, and Kittell's (1957), in which the group which received an intermediate amount of guidance was superior in learning, retention, and transfer to groups receiving either more or less direction. It has been suggested that the "intermediate" amount of guidance provided by Kittell may have actually exceeded the amount Craig provided to his directed group (Ausubel, 1961, p. 52).

One of the few studies that forced learners to discover almost entirely without help provides data in support of the discovery process (Kersh, 1958). The contrasting directed treatment groups were superior in learning rate and immediate recall, but the "no help" group was superior in terms of retention and transfer after a period of approximately one month following the learning period. No evidence was produced to indicate that the no help group understood the rules better. Instead, an explanation was offered in terms of practice. On the basis of a subjective analysis of the subject's comments written on the retests and reported to the experimenter, it was concluded that the learners were motivated to continue the learning process or to continue practicing the task after the formal learning period.

The present experiment was designed to provide formal data concerning the motivating power in question.

## Hypothesis

Each subject had the task of learning the following two rules of addition:

1. *Odd Numbers rule.* The sum of any series of consecutive odd numbers beginning with 1 is equal to the square of the number of figures in the series. (For example, 1, 3, 5, 7, is such a series; there are four numbers, so 4 × 4 is 16, the sum.)

2. *Constant Difference rule.* The sum of any series of numbers in which the difference between the numbers is constant is equal to one-half the product of the number of figures and the sum of the first and last numbers. (For example, 2, 3, 4, 5, is such a series; 2 and 5 are 7; there are four figures, so 4 × 7 is 28; half of 28 is 14 which is the sum.)

The rules can be learned by simple memorization of the task procedure as above. Further, the learner can become cognizant of certain relations

TABLE 1. Number of Subjects (of 10 in Each Cell) Who Used and Stated the Rules Correctly on the Retest

| | Used Rules | | Stated Rules | |
|---|---|---|---|---|
| Treatment Groups | Odd Numbers[a] 1 | Constant Difference 2 | Odd Numbers[a] 3 | Constant Difference[a] 4 |
| Rote Learning: | | | | |
| 3 days | 7 | 7 | 7 | 9 |
| 2 weeks | 7 | 6 | 2 | 6 |
| 6 weeks | 4 | 4 | 0 | 3 |
| Guided Discovery: | | | | |
| 3 days | 6 | 6 | 8 | 9 |
| 2 weeks | 3 | 5 | 3 | 4 |
| 6 weeks | 2 | 3 | 3 | 3 |
| Directed Learning: | | | | |
| 3 days | 4 | 3 | 3 | 4 |
| 2 weeks | 4 | 3 | 1 | 3 |
| 6 weeks | 0 | 3 | 1 | 1 |

[a] Differences between treatment groups and between test periods significant by chi square at or beyond .05 level.

by Li (1957, pp. 416–20). The data included under each of the columns of Table 1 were envisioned as a 2 × 9 contingency table, with 8 *df*. Four separate analyses were then conducted, each of which broke down the chi square into the following components: (*a*) differences between teaching treatments (2 *df*), (*b*) differences between test periods (2 *df*), and (*c*) differences attributed to interaction of treatments and time periods (4 *df*).

None of the interaction effects was significant, indicating that the rate of forgetting did not differ significantly across the teaching treatment groups. A trend analysis of the test data indicated also that the rate of forgetting was constant for all groups (Li, 1957, pp. 226–233).

Otherwise, as pointed out by the footnote references in Table 1, the differences between treatment groups and between test periods were found to be significant for all columns except that headed "Constant Difference 2," for which the observed differences were found not to be reliable.

Perhaps the most striking finding in the present study is that the Rote Learning group was found to be consistently superior in every respect to the other treatment groups. Although this completely unanticipated finding has no direct bearing on the hypothesis in question, it does nevertheless bear clearly upon the related question of "meaningful vs. mechanical" learning. This finding will be discussed in a subsequent section.

Strictly speaking, the hypothesis which the present experiment was designed to test involves only the Guided Discovery and Directed Learning treatments. To support the major hypothesis, the data should have shown that the subjects comprising the Guided Discovery group used the rules after the learning period more frequently than the subjects in the Directed Learning group; and, if so, that the former remembered and transferred the rules more effectively than the latter.

With respect to the frequency of using the rules after the learning period, the results do support the hypothesis. Although the number of subjects in each group who reported that they did use the rules was very small, the difference between the frequency patterns of the two groups in question is statistically significant. Eleven subjects of the 30 in the Guided Discovery group reported that they had used the rules as compared with two subjects in the Directed Learning group. In the Rote Learning group, six subjects of 30 reported in the affirmative.

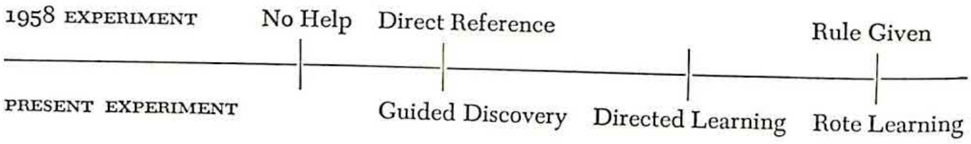
With respect to the relative permanence of the retention and increased transfer effects, the results also support the hypothesis. The Guided Discovery group is clearly superior to the Directed Learning group 3 days after the learning period, and since the rate of forgetting may be presumed to be approximately the same for each treatment group (see statistical analysis above), their initial superiority remains after 6 weeks.

## Discussion

The data from this present experiment do not support the generalization that learning by a process which involves discovery is necessarily superior to learning by more highly directed processes. Indeed, these data suggest that under certain conditions of learning, highly formalized "lecture-drill" techniques, ordinarily considered sterile and meaningless, produce better results than techniques which attempt to develop "understanding."

One explanation for the present results is that they reflect a simple and well known phenomenon, retroactive inhibition. The experimental efforts to inject meaning into the rules amounted to following their initial rote learning with a closely related and complex learning task; thus the Rote Learning group may have surpassed other groups simply because retention among the latter was inhibited by the interpolated learning.

How may the present results be reconciled with those of the previous experiment by Kersh (1958), in which learning by discovery proved markedly superior? The preferred interpretation is that the findings of the two studies are actually complementary. Schematically, the treatments employed in the two experiments may be compared on a line representing the continuum of learning processes: at one extreme, learning without any external direction whatsoever (true self-discovery); at the other, learning by lecture-drill processes (rote learning), as follows:



As is indicated above, the Direct Reference treatment in the 1958 experiment is comparable to the Guided Discovery treatment in the present one; similarly, the Rule Given and Rote Learning groups correspond. The present experiment has no counterpart to the No Help treatment of the previous study; and, in the previous one, the Directed Learning treatment was not represented.

When compared as above, the results of the two experiments are remarkably similar. The initial achievement of the comparable groups in both experiments was very high then dropped to where only about half of each group was able to recall and apply the rules after 4 to 6 weeks. In each experiment the difference in the performance of the Rote Learning and Directed Discovery groups was not notable; if anything, the Rote Learning groups tended to perform slightly better.

With respect to the motivating power of learning by discovery, in the 1958 experiment the superior performance of the No Help subjects on the retest together with their written comments and verbal reports to the experimenter strongly evidenced their increased interest. The present results leave no doubt that there is a tendency for interest to accrue as a result of learning by discovery.

The results of both experiments also are consistent in their failure to support the notion that attempts to provide added meaning will necessarily prolong memory for rules and procedures and will enhance their transfer. On the contrary, both experiments suggest that such attempts may well do more to interfere with learning than enhance it. This does not mean that rote learning is superior to learning with understanding. Rather it means that we need to know much more than we do about meaningful learning and how we come by it.

The relatively poor showing of the Directed Learning group in the present study is partially explained by the subjects' reported failure to practice the rules after the learning period to the extent that the subjects did in other groups. Why the Rote Learning treatment generated more interest than the treatment in question again may reflect nothing more than the original learning was inhibited by the interpolated programed learning. The subjects' unfamiliarity with the instructional procedure may have contributed to their confusion.

Most certainly the data from the two experiments under discussion suggest that the frequently taught principles of learning that pertain to self-discovery and meaning (see introduction) should be restated or qualified. The following statements are offered for further study.

### Learning by Self-discovery

Learning by self-discovery is superior to learning with external direction only insofar as it increases student motivation to pursue the learning task. If sufficiently motivated, the student may then continue the learning process autonomously beyond the formal period of learning. As a result of his added experience, the learner may then raise his level of achievement, remember what he learned longer, and transfer it more effectively. The explanation for the elusive drive generated by independent discovery is not evident, but several have been offered, including the Zeigarnik effect of superior memory for unfinished tasks and the Ovsiankina effect of resumption of incomplete tasks.[1] It also could be explained in terms of operant conditioning; specifically, as a kind of "searching behavior" reinforced by the experimenter's comments and by the subject's own successful progress toward a solution. Whatever the explanation, the motivating power evidently does not appear in strength unless the student is required to learn almost completely without help and expends intensive effort over a period of 15 minutes or more.

### Meaningful Learning

Aside from the advantage the student may come to have academically, he may not benefit from knowing the explanations for rules and procedures he learns, i.e., the pattern of relationships involved. That which is meaningful (understood) may or may not be retained longer and transferred more effectively than that which has been learned by rote. Moreover, superficial efforts to gain understanding after a rule or principle has been memorized may have an inhibitory effect when the student attempts to recall and transfer the original learning. If it is important only that the task be understood (as is most often the case, presumably), the essential relationships may be learned most economically when taught by another person or teaching program, not by process of self-discovery.

## REFERENCES

Ausubel, D. P. Learning by discovery: Rationale and mystique. *Bull. nat. Ass. second. sch. Principals,* 1961, 45, 18–58.

---

[1] The author is particularly indebted to Julius M. Sassenrath and the late Percival M. Symonds for their critical comments and suggestions.

Bruner, J. S. The act of discovery. *Harvard educ. Rev.*, 1961, **31**, 21–32.
Craig, R. C. Directed versus independent discovery of established relations. *J. educ. Psychol.*, 1956, **47**, 223–234.
Kersh, B. Y. The adequacy of "meaning" as an explanation for the superiority of learning by independent discovery. *J. educ. Psychol.*, 1958, **49**, 282–292.
Kittell, J. E. An experimental study of the effect of external direction during learning on transfer and retention of principles. *J. educ. Psychol.*, 1957, **48**, 391–405.
Li, J. C. R. *Introduction to statistical inference*. Ann Arbor, Mich.: Edwards, 1957.

# 33

# Some Motivational Factors in Cooperation and Competition — *Marvin E. Shaw*

*The author indicates that cooperation and competition involve at least two factors: motivation and procedure. Two experiments are reported—one involving a perceptual-motor task and the other a memory-reasoning task—in which it was attempted to isolate motivational factors in both cooperation and competition. On the basis of his findings the author concludes that competitive situations arouse stronger motivation to achieve than do cooperative ones. However, this stronger motivation can result in poorer performance if disruptive responses are introduced by the energizing component of motivation. The author's findings point to some interesting inferences about the motivational component of cooperation and competition in classroom situations.*

THE CONCEPTS of cooperation and competition have been used to refer to a wide variety of situations (*1, 3, 12, 13, 15*). Without raising the issue of what these terms "really mean," they are used in this report to refer to behavior directed toward the same goal by at least two individuals. In the cooperative situation the movement of one member of a group toward the goal will to some extent facilitate the movement of other members toward the goal; in the competitive situation the movement of one member toward the goal will to some extent impede the movement of other members

Reprinted from *Journal of Personality*, 1958, **26**, 155–169, by permission of the author and the Duke University Press.

toward the goal. These definitions are essentially those given by Deutsch (3).

Although much has been written concerning cooperation and competition, relatively little experimental work has been done to isolate the variables operating in these two processes. May and Doob (13) have summarized the literature up to 1937 and report no more than a dozen laboratory studies dealing with the effects of cooperation and competition. These were not directed primarily toward comparing the relative efficiency of the two situations, but rather with comparing these with the individual situation.[1] As May and Doob pointed out, the experimental evidence is ambiguous and apparently contradictory in respect to the effects of cooperation and competition upon performance. "Sometimes both quality and quantity were increased, sometimes one was increased and the other diminished" (13, p. 155).

Since the May and Doob survey only a few attempts have been made to compare the relative effectiveness of cooperation and competition. Deutsch (4) has presented evidence that under certain conditions greater group or organizational productivity will result when the members are cooperative rather than competitive in their interrelationships. Blau (2) reported that comparison of two groups of interviewers in a public employment agency reveals that whereas the more competitive group is less productive, the more competitive individuals in the competitive group are relatively more productive than the more cooperative individuals. Smith, Madden, and Sobol (16) observed groups discussing a case study under cooperative and competitive orientations, and later tested for recall of the discussion material. They found that during the discussion period the cooperative groups were more productive of ideas than were the competitive groups, but there was no significant difference between the two conditions in frequency of recall of the discussion material.

There are several difficulties in evaluating the results of studies like those presented above. Aside from the fact that different investigators have studied groups and individuals whose tasks differed considerably, they have failed to isolate factors contributing to productivity. Cooperation and competition involve at least two aspects: (*a*) the motivational factor, or the individual's need to achieve under the two conditions, and (*b*) the procedural factor, or the relative effectiveness of the two social arrangements in the attainment of goals (i.e., simply because of the nature of the task one procedure or the other may be the more effective one). In addition, it is questionable that the conditions necessary for comparing the relative effectiveness of cooperation and competition were established in the experiments discussed in the preceding paragraph. For example, cooperative groups (individuals working together toward a common goal) usually have been motivated to perform the assigned task by causing them to compete with other cooperative groups. Thus, the performance of competitive groups (individuals competing among themselves for a common goal) has been compared with the performance of cooperative groups competing among themselves for a common goal.

---

[1] In fact, most investigations up to 1937 were concerned with problems of defining conditions which favored the development of either cooperative or competitive behavior. May and Doob suggested 68 problems for future research; only three of these were concerned with performance under conditions of cooperation and competition.

The experiments reported in this paper attempted to isolate the motivational factors in cooperation and competition by using a task which is intrinsically interesting to Ss and by making performance dependent upon S's own efforts regardless of whether he believed himself to be in a cooperative or in a competitive situation.

## Experiment I

### Method

*Experimental task.* In order to test the ideas presented above a task was needed which had the following characteristics: (*a*) it should be inherently interesting to most Ss so that external incentives need not be imposed to keep them working on the task; (*b*) it should permit precise and objective evaluation of individual performance; (*c*) it should be a task which two or more persons could attack simultaneously. Unfortunately, tasks which satisfy the first two of these criteria seldom satisfy the third, and vice versa, so it was necessary to select one which met the first two and then to modify it to satisfy the third requirement also. The tracking task seemed to be a desirable one from this standpoint, especially with respect to task interest. The description of the apparatus which follows shows the modifications which were introduced to permit two persons to work at the task simultaneously.

*Apparatus.* The characteristics of apparatus typically used in tracking experiments have been described in detail many times (e.g., *9*, *10*, *11*). The essential components are a rotating target, a rotating cursor, and a handwheel which controls the movement of the cursor. The task is to keep the cursor aligned with the target, which follows a course characterized by variable speed and many reversals of direction. A second target and a second handwheel were added to the apparatus used in this experiment. Both targets were mounted on a rotating wheel whose axis was the same as that of the cursor. One target was painted white and the other red. The white target was used for scoring purposes. Distance between targets, measured from tip to tip, was two inches. Target velocity and direction were controlled by a cam and ball disc integrator.

The cursor could be controlled by either of two handwheels, but not by both simultaneously. Magnetic clutches, secretly operated by *E*, determined which handwheel controlled the cursor at any given time. The cursor made one complete revolution for every five revolutions of the handwheel.

There were two devices for measuring performance. One indicated the amount of time the cursor was actually in the target area during each trial. The target area was actually a copper strip 1½ inches long, but the target itself, as viewed by S, was the same width (⅛ inch) as the cursor. When the cursor was in the target area it was in contact with the copper strip; an appropriate electrical circuit permitted a timer to run as long as contact was maintained. The other measuring device gave the average distance between cursor and target during a trial. A potentiometer with its body connected to the target shaft and its arm connected to the cursor shaft provided error information to one of two integrators, one being for lead errors and the other for lag errors. The integrators had a full scale value of ten degrees. This means that if an S's average error for one minute was ten degrees, the integrator indicator would move through its entire scale, which in this case is five units. The two measures of tracking accuracy tend to be correlated, but are not identical; it is possible for two Ss to earn identical scores according to one measure but differ greatly according to the other.

*Procedure.* The procedure was designed to create three experimental conditions: (*a*) a condition in which S perceived himself to be in a cooperative situation; (*b*) a condition in which S perceived himself to be in a competitive situation; and (*c*) a condition in which S worked alone. The first two conditions were established with the help of a confederate who played the role of a naïve S. Ten Ss were run in each condition, half of which were male and half female in each case.

Except for the variations necessary to create the three experimental conditions, the procedure was essentially the same for all Ss. Immediately after entering the experimental room Ss were seated before the tracking apparatus, with the naïve S always occupying the same position. The apparatus was explained, and each person was given a practice trial during which his crank controlled the cursor. Following this, instructions designed to fit the condition were given. The cooperative Ss were instructed as follows:

> Your task will be to keep the cursor lined up with the white target, working together as a pair. Your score on each trial will be total time on target, regardless of who is actually controlling the cursor at any given time. In other words, this is a cooperative situation; whenever one person is succeeding, the other is succeeding also.

In the competitive situation, Ss were told:

> Your task (indicating naïve S) will be to keep the cursor lined up with the white target, whereas your task (indicating confederate) will be to keep the cursor lined up with the red target. Your score on each trial will be the total time the cursor is lined up with your own individual target. In other words, this is a competitive situation; whenever one person is succeeding, the other is necessarily failing.

In both the cooperative and the competitive situation, Ss were led to believe that by means of an "electronic device" both handcranks influenced the movement of the cursor simultaneously. Thus it appeared to the cooperative S that a good score could be earned if either person performed well; it appeared to the competitive S that a good score could be earned only if he were able to exert more influence on the cursor than his opponent. It may be worth noting that physical strength was not a factor; Ss could not determine from kinesthetic cues whether the other crank was being operated.

Ss in the alone situation were told simply that the task was to keep the cursor lined up with the white target and that the score on each trial would be total time on target.

After answering questions pertaining to procedure, Ss were given 22 trials of 30 sec. each. Trials 1, 2, 3, 6, 10, 15, and 19 were not scored, but were introduced to help maintain the illusion that the confederate was a naïve S. During these trials, *E* randomly switched control of the cursor from one handcrank to the other. Performance scores were recorded for 15 sec. of each of the remaining 15 trials, during which time S had complete and exclusive control of the cursor. During the other 15 sec. of each trial, control was again shifted randomly from one handcrank to the other. There was no evidence that any S suspected that the situation was different from that described by the instructions.

At the end of the session, each S was asked to complete a questionnaire which required ratings of task interest and task satisfaction, and indications of relative preference for competing, cooperating, and working alone.[2] Then *E* explained the nature of the deception and the reason for it.

**Results**

The results are summarized in Table 1, and the outcomes of analyses of variance of time on target and integrated error scores are given in Table 2. Although Bartlett's test for homogeneity of variance did not require the rejection of the hypothesis of random sampling from a population with a common variance, the means and variances of time and error scores tended to be correlated. Therefore, all scores were converted by the square root transformation before the analyses of variance were performed.

As can be seen in Tables 1 and 2, the experimental conditions produced markedly different results. The cooperative situation was the most effective condition both with respect to time on target and integrated error, whereas the competitive situation was least effective according to both measures, with the individual situation falling between these two.

---

[2] These comparisons were made with respect to the general case, rather than with respect to the particular experimental situation. This was deemed appropriate since all Ss did not experience all experimental conditions.

TABLE 1. Means and Standard Deviations of Accuracy Scores and Ratings as a Function of Experimental Conditions in Experiment I

| | Males | | | Females | | |
|---|---|---|---|---|---|---|
| | Coop | Comp | Ind | Coop | Comp | Ind |
| Time on Target (Secs.) | | | | | | |
| Mean | 10.11 | 6.75 | 9.45 | 8.60 | 6.86 | 7.40 |
| SD | 0.93 | 2.52 | 1.35 | 1.47 | 1.53 | 2.08 |
| Integrated Error (arbitrary scale) | | | | | | |
| Mean | 2.64 | 5.66 | 3.13 | 3.76 | 5.75 | 4.97 |
| SD | 1.07 | 2.81 | 1.22 | 1.43 | 2.28 | 2.93 |
| Ratings of Satisfaction | | | | | | |
| Mean | 7.0 | 7.3 | 7.0 | 5.5 | 7.9 | 5.5 |
| SD | 1.4 | 1.6 | 1.4 | 0.9 | 0.9 | 0.5 |
| Ratings of Task Interest | | | | | | |
| Mean | 8.0 | 7.8 | 7.4 | 5.6 | 7.3 | 7.0 |
| SD | 1.4 | 1.1 | 0.7 | 1.7 | 1.6 | 0.5 |

TABLE 2. Summary of Analyses of Variance Based on Transformed Time and Error Scores from Experiment I

| | | Time on Target | | | Integrated Error | | |
|---|---|---|---|---|---|---|---|
| Source | *df* | Mean Square | *F* | *p* | Mean Square | *F* | *p* |
| Motivation (M) | 2 | 9.215 | 5.91 | $<.01$ | 11.935 | 4.45 | $<.05$ |
| Sex (S) | 1 | 3.910 | 2.51 | | 6.550 | 2.45 | |
| Trials (T) | 14 | 0.485[a] | 7.35 | $<.01$ | 0.663[a] | 4.87 | $<.01$ |
| Ss Within | 24 | 1.559 | | | 2.678 | | |
| M × S | 2 | 1.965 | 1.26 | | 1.485 | | |
| M × T | 28 | 0.057[a] | 0.86 | | 0.113[a] | 0.83 | |
| S × T | 14 | 0.072[a] | 1.09 | | 0.202[a] | 1.49 | |
| M × S × T | 28 | 0.100[a] | 1.51 | | 0.115[a] | 0.85 | |
| Residual error | 336 | 0.066 | | | 0.136 | | |
| Total | 449 | | | | | | |

[a] Tested against residual error.

Figure 1 shows the effects of practice on the time on target scores for the three experimental conditions. Despite some notable inversions, performance improved rather steadily as a function of practice. It seems safe to interpret the significant trials term in the analysis of variance as indicating that learning occurred in all conditions. Similar results were found for error scores.

Since the three conditions differed on the first trial, it might appear that differences among the motivational conditions were due to differences in initial skill. However, this hypothesis was rejected because the three groups were not significantly different on the initial 1 minute practice trial (Means were 30.2, 29.7, and 31.0 seconds for the competitive, cooperative, and control groups, respectively).

The ratings of satisfaction, however, reveal a different pattern. The competitive situation was found to be more satisfying ($F = 4.41$, $p < .05$) than either the cooperative or the individual situation. Also, Ss indicated that they preferred a competitive situation to either the cooperative or the individual situation ($\chi^2 = 8.50$, $p < .01$).

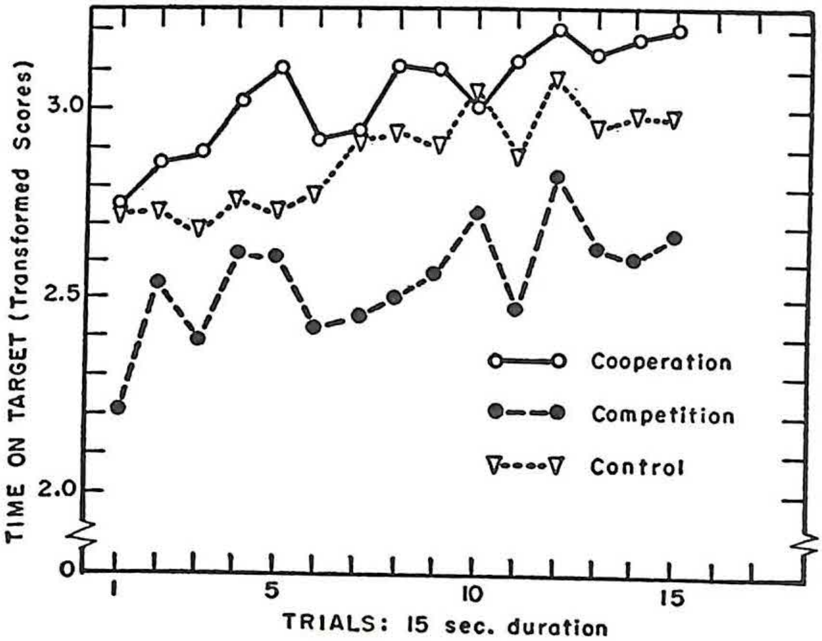


FIG. 1. *Mean time on target as a function of practice in the three motivational conditions.*

There were no reliable differences between experimental conditions with respect to ratings of task interest, but female Ss found the task less interesting than did the male Ss ($F = 6.13$, $p < .05$). This was the only sex differences found.

It was also possible to examine the consistency of performance from trial to trial for a given S. This analysis revealed that the competitive Ss tended to be less consistent than either the cooperative Ss or those working alone; however, this difference was not statistically reliable.

## Discussion

The results presented above show clearly that under these experimental conditions performance was better when Ss believed themselves to be in a cooperative situation rather than in a competitive one, whereas individual performance was somewhere between these two. One might assume that this finding demonstrates that cooperation arouses stronger motivation to achieve than does either competition or solitary activity. If this hypothesis is true, however, it seems odd that *Ss* in the cooperative situation should not also enjoy the task more than Ss in either of the other two conditions. It may be assumed that among college students, challenging (motivating) situations are more satisfying than are less challenging situations, especially when performance on the challenging task is superior to that on the less challenging task. Likewise, one might expect a general preference for cooperative situations if they typically arouse the stronger motivation.

Other evidence also leads to the questioning of the hypothesis that better performance means stronger motivation. Subjective observations by *E* indicated that Ss in the competitive situations actually tried harder than did those in either the cooperative or the individual situation. For example, competitive Ss quite evidently cranked faster, frequently causing the apparatus to "dance" about, showed more concern about how they were performing, etc. There were also frequent signs of emotion in the competitive situation, such as giggling, reddening, and irrelevant comments, which were much less frequent in the individual situation and almost entirely absent in the cooperative situation.

An alternative explanation for the obtained results might be that competition aroused a stronger motivation to achieve which, rather than improving S's performance, actually interfered with it. That is, on tasks such as the present one, which utilizes a perceptual-motor skill, trying too hard may be detrimental to effective performance. Assuming that competition aroused the stronger motivation, then the energizing component of motivation should cause Ss to crank faster when trying to realign the cursor and target. This behavior, however, leads to overcorrections and hence decreased accuracy. This seems to have happened to the competitive Ss in this experiment and, to a lesser extent, to Ss who worked alone.

In addition, emotional factors are involved in several ways. In the competitive and individual situations, responsibility for a poor performance could not be shifted easily to another person. This probably resulted in a condition which was threatening to S's self-esteem. In the competitive situation there existed an additional threat in the form of the competitor. Thus it is likely that Ss found the competitive situation relatively more stressful than did the individual Ss, who in turn experienced more stress than did the cooperative Ss. As has been demonstrated in a number of experiments (*6, 14, 17, 19*), stress (defined as a threat to the attainment of some goal, such as threat of failure, threat to self-esteem, etc.) usually produces a decrement in performance as measured by accuracy scores.[3] The results of this experiment are adequately understood by assuming that the relatively stronger motivation in the competitive and individual conditions led to increased efforts but less than perfect performance, thus producing frustration. Frustration initiated task-irrelevant (inappropriate) responses which interfered with the task-relevant responses, thus leading to the relatively poor performance.

The evidence supporting the above interpretation is largely subjective, but nevertheless fits in with the notion that the competitive Ss performed less well, not because they were less motivated, but because they were trying too hard. This interpretation is also consistent with the ratings of satisfaction and the preference for a competitive situation, as well as with the greater between-trials variance in the competitive as compared with the other two conditions.

The interference hypothesis outlined above is based in part upon the nature of the task which is particularly susceptible to the disruptive influences associated with the energizing component of motivation. The task also involved operations novel to most Ss and hence may have been threatening to self-esteem. It was reasoned that if the task were such that the energizing component of motivation was less disruptive, and if the task involved processes familiar to Ss, the observed differences among experimental conditions might be considerably different. For example, in the individual situation the effects of the energizing component should be eliminated and the threatening aspect should be at least decreased; therefore, performance should improve relative to the other conditions. The same should be true for the competitive condition, but to a lesser degree since the "threatening" other person would still be present. The purpose of the following experiment was to test these expectations.

## Experiment II

### Method

*Experimental task and apparatus.* The task chosen for the second experiment can best be explained by describing the apparatus, which is a drastically modified version of the Yerkes (*18*) multiple choice apparatus. It consisted of a panel upon which a light and four button type switches had been mounted, and a separate control box which allowed *E* to select which of the four switches would turn on the light when pushed. The switches were numbered from one through four. A given problem consisted of a sequence of settings, such as 4, 3, 2, 1, which was repeated over and over. The task was to discover this sequence. Some reasoning was involved, but it was largely a memory task.

---

[3] Some experiments have shown that this effect is dependent upon individual characteristics (*5, 7, 8*), but it seems likely that individuals showing an increment in performance as a function of stress will be in the minority in any randomly chosen sample (*6*).

*Procedure.* The procedure was again designed to create three experimental conditions in which the naïve S perceived himself as either cooperating, competing, or working alone. In the cooperative and competitive situations a confederate played the role of a naïve S, but actually contributed nothing to the variance of performance among Ss. Ten Ss were run in each of the three conditions.

Except for the variations necessary to create the three experimental conditions, the procedure was similar for all Ss. The nature of the task was explained, and when *E* was sure that S understood the problem, the following instructions were given in the cooperative and competitive conditions:

> This will be the procedure. When I say "Go," one of you (indicating the naïve S) select a button and press it. If the light comes on, that ends the trial; if it does not come on, the other person (indicating the confederate) may select a button and press it. Regardless of what happens this time, the trial ends. Thus at the end of Trial 1 you should either know which is the correct switch, or that it is one of the two which were not pressed. This process will be repeated until you can name the sequence or until twenty trials have been completed.

The instructions in the individual condition were the same except that the second sentence was changed to read, "If the light comes on, that ends the trial; if it does not come on, you will be allowed to push one additional button which I will select." In selecting which button to press, both *E* and the confederate followed a system which made the selection contingent upon S's choice. On odd numbered problems, if S chose an odd numbered switch, then the confederate (or *E*) also chose an odd numbered switch; if he chose an even number, then the confederate (or *E*) chose an even number. On even numbered problems, the selection was always opposite to S's choice. This successfully prevented S from becoming suspicious of the confederate, and assured that he did not help Ss in one condition more than in others.

It can be seen that S could learn the sequence in no more than eight trials regardless of what the confederate or *E* chose to do, providing they did not simply choose the *same* switch as S on a given trial. On the first trial for each setting in the four digit sequence, he could narrow the possibilities to two, and on the second time through he could decide between these two alternatives. This assumes, of course, that S sees this simple logic and that he remembers what happened on previous trials, since Ss were not allowed to keep written records. However, S might learn the sequence in less than eight trials if any correct switch happened to be selected during the first four trials. Thus, there did exist some possibility that S would get some fortuitous help from the confederate or from *E*. Records were kept of the number of times the correct switch was selected by anyone other than S, but it was found that this number did not differ significantly for the different conditions, either when the total number of "hits" was used as the measure or when the measure was the number of hits during the first four trials. Therefore, differences among experimental conditions could not be attributed to differences in amount of help given by the confederate or by *E*.

To establish the competitive situation, Ss were told that their score would be determined by the number of tasks completed before the other person and also by the difference between Ss in the number of trials required to discover the sequence. In other words, it was to his advantage to prevent the other person from discovering the sequence even after he himself had learned it.[6]

For the cooperative situation, Ss were told that their score would be the number of trials required to discover the sequence regardless of which person actually hit upon the correct soluition. In this condition, the confederate merely agreed with any answer suggested by S, usually saying, "Let's try it and see." Occasionally, the confederate would see that S was writing down the correct answer and would write it on his own card at the same time.

In the individual situation, Ss were told that their score would be the number of trials required to learn the sequence.

All answers were written on 3 × 5 cards which were then handed to *E*. S (or the confederate) could submit an answer card at any time during the test. If the answer submitted was incorrect, *E* said so, and the test continued until a correct answer was submitted or until 20 trials had been completed. The frequency of incorrect answers was very low in all three conditions, but the total was slightly higher in the competitive than in the other two conditions.

Each S solved ten problems. Order of presentation was randomized within a given condition but was standardized across conditions. After all problems had been completed, S was required to rate his enjoyment of the task, his performance on the task, his motivation while working at the task, and the amount of effort expended on each problem. After this questionnaire had been completed, S was questioned regarding his attitude toward the task, how he felt during the experiment, etc. Then the purpose of the experiment was explained, and he was requested to refrain from talking to others about his experience.

---

[6] It is conceivable that S might try to confuse his opponent by making irrational choices before having learned the solution himself, but there was no evidence that any S employed this stratagem.

### Results

Table 3 gives the results of the second experiment, and the outcome of the analysis of variance of "trials to solve" is given in Table 4. As can be seen by comparing these data with the results presented in Tables 1 and 2, the outcomes of the two experiments are quite similar. The competive Ss performed less effectively, as measured by trials to solve, but indicated that they enjoyed the task more than did either the cooperative or the individual Ss ($F = 4.08$; $df$ 2, 27; $p < .05$). Also, they rated amount of effort expended higher ($F = 3.62$; $df$ 2, 27; $p < .05$) than did Ss in the other two conditions. Although differences were not statistically reliable, competitive Ss also rated their motivation higher than did Ss in the other conditions. However, differences in performance between the individual and the cooperative situations were not found, despite the fact that ratings of effort were greater in the individual situation.

TABLE 3. Means and Standard Deviations of Trials to Solve and Ratings as a Function of Experimental Conditions in Experiment II

| | Coop | Comp | Ind |
|---|---|---|---|
| Trials to Solve | | | |
| Mean | 9.9 | 12.0 | 10.1 |
| SD | 1.4 | 1.5 | 1.1 |
| Ratings of Satisfaction | | | |
| Mean | 4.1 | 4.6 | 3.6 |
| SD | 0.4 | 0.7 | 1.0 |
| Ratings of Motivation | | | |
| Mean | 3.8 | 4.0 | 3.9 |
| SD | 0.5 | 0.8 | 0.9 |
| Ratings of Effort | | | |
| Mean | 3.5 | 4.1 | 3.8 |
| SD | 1.1 | 0.7 | 1.0 |

TABLE 4. Summary of the Analysis of Variance of Trials to Solve—Experiment II

| Source | $df$ | Mean Square | $F$ | $p$ |
|---|---|---|---|---|
| Motivation (M) | 2 | 116.56 | 5.79 | <.01 |
| Trials (T) | 9 | 46.22[a] | 2.38 | <.05 |
| Ss Within | 27 | 20.13 | | |
| M × T | 18 | 12.02 | | |
| Residual error | 243 | 19.44 | | |
| Total | 299 | | | |

[a] Tested against residual error. (Ss Within x Trials.)

As shown in Table 4, performance scores varied significantly as a function of trials (practice). As in the first experiment, performance improved as a function of practice, so we may assume that learning occurred.

In the interview period which followed the administration of the questionnaire, Ss in the competitive condition reported that they definitely felt "on trial" and that an inferior performance relative to the other person would reflect unfavorably upon their own ability. Those in the individual condition said that they felt their ability was being tested but felt no particular pressure or threat to self-esteem. The cooperative Ss reported complete relaxation; they apparently felt that the other person could be relied upon to perform well or to share the blame for a poor performance.

### Discussion

The results of the second experiment agree with the findings reported in Experiment I with respect to the relative effectiveness of cooperative and competitive situations. In both experiments the cooperative Ss performed better than did the competitive Ss, despite the fact that the two tasks were quite dissimilar. The results also agree reasonably well with findings reported by others (*4, 16*) who did not attempt to separate motivational and procedural factors. This finding thus appears to have some generality.

The meaning of these findings in regard to the motivational factor is less clear. The fact that with the memory-reasoning task there were no reliable differences between the individual and the cooperative situations is in accord with the interference hypothesis. Ss in the individual situation were apparently somewhat more highly motivated than the cooperative Ss (as indicated by the higher ratings of effort), but at the same time experienced some threat (since they were still solely responsible for the performance score) which produced some interfering responses. The two forces operating in opposite directions resulted in performance similar to that of the cooperative Ss.

Although comparisons across tasks are extremely dubious, it does seem that differences between the competitive and cooperative conditions were greater on the perceptual-motor task than on the memory-reasoning task (a mean difference of 37 per cent on the perceptual-motor task as compared with a mean difference of 21 per cent on the memory-reasoning task). If such a comparison is meaningful, this difference also agrees with expectations based upon the interference hypothesis.

In conclusion, it is believed that competitive situations arouse stronger motivation to achieve than do cooperative situations; this stronger motivation, however, results in poorer performance as measured by accuracy scores which are particularly susceptible to disruptive responses introduced by the energizing component of motivation and to interference by task-irrelevant responses resulting from threat to self-esteem. Measures of motor output, such as speed scores, probably would indicate a superior performance by Ss in the competitive situation relative to that of Ss in the cooperative situation.

## Summary

Cooperation and competition involve, at the least, both motivational and procedural factors. Two experiments were conducted to test the relative effectiveness of cooperation and competition under conditions which were controlled for the effects of the procedural variable. One experiment employed a perceptual-motor task and the other a memory-reasoning task.

The results of the two experiments agreed in finding the cooperative situation to be more efficient, but less satisfying, than the competitive situation. The two sets of results disagreed in that in the first experiment the individual situation was less efficient than the cooperative situation, whereas in the second experiment this difference was not found.

The reasons for these findings are not clear, but an interpretation was offered in terms of the disruptive effects of the competitive situation via interfering responses introduced by the energizing component of motivation. It was also suggested that some interference may have occurred in the individual and competitive situations as a result of task-irrelevant responses due to threat to self-esteem.

It was pointed out that although there is still some question about the motivational effects of cooperation and competition, the data show that the relative superiority of the cooperative situation does not depend entirely upon procedural factors.

## REFERENCES

1. Barnard, C. I. *The functions of the executive.* Cambridge, Mass.: Harvard Univer. Press, 1938.
2. Blau, P. M. Co-operation and competition in a bureaucracy. *Amer. J. Sociol.*, 1954, **59**, 530–535.
3. Deutsch, M. A theory of co-operation and competition. *Hum. Relat.*, 1949, **2**, 129–152.
4. Deutsch, M. An experimental study of the effects of co-operation and competition upon group process. *Hum. Relat.*, 1949, **2**, 199–232.
5. Hurlock, E. B. The value of praise and reproof as incentives for children. *Arch. Psychol.*, 1924, **11**, No. 71.
6. Lazarus, R. S., Deese, J., & Osler, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.
7. Lazarus, R. S., & Eriksen, C. W. Psychological stress and its personality cor-

relates: Part I, The effects of failure stress upon skilled performance. *J. exp. Psychol.*, 1952, 43, 100–105.

8. Lindsley, D. B. *A study of performance under speed stress.* OSRD, 1946, Publ. Bldg. No. 1838. Washington: U.S. Department of Commerce, 1946.
9. Lincoln, R. S. Visual tracking: III. The instrumental dimension of motion in relation to tracking accuracy. *J. appl. Psychol.*, 1953, 37, 489–493.
10. Lincoln, R. S., & Smith, K. U. Transfer of training in tracking performance at different target speeds. *J. appl. Psychol.*, 1951, 35, 358–362.
11. Lincoln, R. S., & Smith, K. U. Visual tracking: II. Effects of brightness and width of target. *J. appl. Psychol.*, 1952, **36**, 417–421.
12. Maller, J. B. Cooperation and competition: An experimental study in motivation. *Teach. Coll., Contr. to Educ.*, No. 384, 1929.
13. May, M. A., & Doob, L. W. Co-operation and competition. *Soc. Sci. Res. Council.*, 25, 1937.
14. McClelland, D. C., & Apicella, F. S. Reminiscence following experimentally induced failure. *J. exp. Psychol.*, 1947, 37, 159–169.
15. Mead, M. *Co-operation and competition among primitive peoples.* New York: McGraw-Hill, 1937.
16. Smith, A. J., Madden, H. E., & Sobol, R. Productivity and recall in cooperative and competitive discussion groups. *J. Psychol.*, 1957, 43, 193–204.
17. Verville, E. The effect of emotional and motivational sets on the perception of incomplete pictures. *J. genet. Psychol.*, 1946, **69**, 133–145.
18. Yerkes, R. M. A new method of studying the ideational behavior of mental defective and deranged as compared with normal individuals. *J. comp. Psychol.*, 1921, **1**, 369–394.
19. Zeller, A. F. An experimental analogue of repression: II. The effect of individual failure and success on memory measured by relearning. *J. exp. Psychol.*, 1950, **40**, 411–422.

# 34

# The Use of Advance Organizers in the Learning and Retention of Meaningful Verbal Material

## — *David P. Ausubel*

*The author assumes that "cognitive structure is hierarchically organized in terms of highly inclusive concepts under which are subsumed less inclusive subconcepts and informational data." Based on subsumption theory, this research tests the hypothesis that both the learning and retention of meaningful but unfamiliar verbal material can be facilitated by introducing it with an "advance organizer," i.e., an introductory passage containing relevant subsuming concepts. Results of the experiment support the hypothesis. It is suggested that greater use of appropriate advance organizers in the teaching of meaningful verbal material could lead to more effective retention.*

THE PURPOSE of this study is to test the hypothesis that the learning and retention of unfamiliar but meaningful verbal material can be facilitated by the advance introduction of relevant subsuming concepts (organizers). This hypothesis is based on the assumption that cognitive structure is hierarchically organized in terms of highly inclusive concepts under which are subsumed less inclusive subconcepts and informational data (Ausubel, Robbins, & Blake, 1957). If this organizational principle of progressive differentiation of an internalized sphere of knowledge does

Reprinted from *Journal of Educational Psychology,* 1960, **51,** 267–272, by permission of the author and the American Psychological Association.

in fact prevail, it is reasonable to suppose that new meaningful material becomes incorporated into cognitive structure in so far as it is subsumable under relevant existing concepts. It follows, therefore, that the availability in cognitive structure of appropriate and stable subsumers should enhance the incorporability of such material. If it is also true that "meaningful forgetting" reflects a process of memorial reduction, in which the identity of new learning material is assimilated by the more inclusive meaning of its subsumers (Ausubel et al., 1957), the same availability should also enhance retention by decelerating the rate of obliterative subsumption.

In the present study, appropriate and relevant subsuming concepts (organizers) are deliberately introduced prior to the learning of unfamiliar academic material, in order to ascertain whether learning and retention are enhanced thereby in accordance with the theoretical premises advanced above.

## Method

### Subjects

The experimental population consisted of 120 senior undergraduate students (78 women and 32 men) in four sections of an educational psychology course at the University of Illinois. All Ss were enrolled in one of eight teacher education curricula at the secondary school level. Students specializing in industrial education and in vocational agriculture were excluded from the study since they had received specific instruction in the topic covered by the learning passage. The experiment was conducted separately in each section as a required laboratory exercise and was performed during regularly scheduled class hours. In order to maximize ego-involvement, Ss were informed that after the data were processed their individual scores, as well as the class results, would be reported to them.

### Learning Passage and Test of Retention

The learning material used in this study was a specially prepared 2,500-word passage[1] dealing with the metallurgical properties of plain carbon steel. Emphasis was placed on such basic principles as the relationship between metallic grain structure, on the one hand, and temperature, carbon content, and rate of cooling, on the other. Important factual information (e.g., critical temperatures), however, was also included, and basic principles were also applied to such technological processes as heat treatment and tempering.

The metallurgical topic was chosen on the basis of being generally unfamiliar to undergraduates in liberal arts and sciences (i.e., not ordinarily included in chemistry courses), but still sufficiently elementary to be both comprehensible and interesting to novices with no prior background in the field. The criterion of unfamiliarity was especially crucial because the purpose of the study was to ascertain whether advance organizers could facilitate retention in areas of knowledge *new* to learners. By using unfamiliar material it was also possible to ensure that all Ss started from approximately the same baseline in learning the material. Empirical proof of unfamiliarity was sought, therefore, by administering the retention test on the steel passage to a comparable group of naive Ss who had *not* studied the material; but although this latter group of Ss made scores which, on the average, were only slightly and not significantly better than chance, it was evident from later analysis of the experimental data that scores earned by Ss who *had* studied the passage were related to both sex and field of specialization. Male students and majors in science and art were better able to learn and retain the steel material than were female students and majors in English, foreign languages, music, and the social sciences. Hence, the criterion of unfamiliarity was not completely satisfied, in as much as these differences undoubtedly reflected, in part, variability in relevant incidental experience influencing the learnability of the material.

---

[1] Appreciation is expressed to Robert M. Tomlinson for assistance in the preparation of the learning passage.

Knowledge of the steel passage was tested by a 36-item multiple-choice examination with a corrected split-half reliability of .79. Test questions covered principles, facts, and applications, and were selected by an item analysis procedure from a larger population of items. Scores on the test showed a satisfactory range of variability and were distributed normally. Since it was intended as a power test, no time limit was imposed.

**Procedure**

It was first necessary to equate experimental and control groups on the basis of ability to learn an unfamiliar scientific passage of comparable difficulty. The passage used for this purpose was concerned with the endocrinology of human pubescence and was approximately 1,800 words long. Ss were given 20 minutes to read and study this material, and were tested immediately thereafter by a 26-item multiple-choice test with a corrected split-half reliability of .78. (The unfamiliarity of the material has been previously ascertained by administering the test to a comparable group of naive Ss who had not studied the passage, and obtaining a mean score only slightly and not significantly greater than chance.) Test scores on the pubescence passage were normally distributed and correlated .64 on a product-moment basis with test scores on the steel passage. *F* tests were performed on the variance ratios of the pubescence material test scores for all possible combinations of the four sections, and none approached significance at the .05 level of confidence. It was considered justifiable, therefore, to treat the retention scores of experimental and control groups on the steel passage as if derived, respectively, from one large class rather than from four separate sections.

Ss in each of the four sections were matched on the basis of test scores on the pubescence material and assigned to experimental and control groups. Experimental and control treatments were then administered simultaneously to experimental and control Ss, respectively, within each section. This procedure was possible because the two treatments consisted of studying identical appearing introductory passages differing only in content. The use of this procedure also provided the important methodological advantage of holding instructor, class, and situational variables constant for both groups. Each introductory passage of approximately 500 words was studied twice, 5 minutes each time, by the appropriate group of Ss. The two occasions were 48 hours and immediately before exposure to the main learning passage.

The experimental introductory passage contained background material for the learning passage which was presented at a much higher level of abstraction, generality, and inclusiveness than the latter passage itself. It was designed to serve as an organizing or anchoring focus for the steel material and to relate it to existing cognitive structure. Principal emphasis was placed, therefore, on the major similarities and differences between metals and alloys, their respective advantages and limitations, and the reasons for making and using alloys. Although this passage provided Ss in the experimental group with relevant background concepts of a general nature, it was carefully designed not to contain specific information that would confer a direct advantage in answering any of the questions on the steel test. This latter criterion was tested empirically and shown to be warranted when a comparable group of Ss made only a slightly better than chance mean score on the steel test after studying the introductory passage alone.

The control introductory passage, on the other hand, consisted of such historically relevant background material as the historical evolution of the methods used in processing iron and steel. This type of introductory material is traditionally included in most textbooks on metallurgy and is presumably intended to enhance student interest. In contrast to the introductory passage given to the experimental group, it contained no conceptual material that could serve as an ideational framework for organizing the particular substantive body of more detailed ideas, facts, and relationships in the learning passage.

It was methodologically necessary to provide this control treatment in order that any obtained difference between experimental and control groups could be attributed to the particular nature of the experimental introductory passage (i.e., to its organizing properties) rather than to its presence per se.

Both groups studied the steel passage for 35 minutes and took the multiple-choice steel test 3 days later. Since it was evident from a comparison of test scores on the steel and pubescence passages that scores on the steel test were related to Ss' sex and major field, it was necessary to hold these latter factors (as well as pubescence test scores) constant. Hence, it was no longer possible to use the originally matched pairs of Ss within each section. Sufficient Ss were also not available to rematch individual pairs of Ss on all three variables. By matching experimental and control Ss across sections, however, it was possible to equate two groups of 40 Ss each for sex, pubescence scores, and field of specialization. The crossing of sectional lines in this rematching procedure was justifiable in view of the intersectional homogeneity of variance.

### Results and Discussion

The distribution of steel test scores for both experimental and control groups did not deviate significantly from the normal.[2] The mean steel test score of the experimental group was 16.7, as compared to 14.1 for the control group and a mean chance score of 7.2 (one-fifth of 36). The standard deviations of the two groups were 5.8 and 5.4, respectively. The difference between the means[3] of the experimental and control groups was almost significant at the .01 level for a one-tailed test.

TABLE 1. Retention Test Scores of Experimental and Control Groups on Learning Passage

| Group | Type of Introduction | Mean | *SD* |
|---|---|---|---|
| Experimental | Substantive | 16.7 | 5.8 |
| Control | Historical | 14.1 | 5.4 |

NOTE: Chance Score on the multiple-choice test of 36 items is 7.2. The difference between the means in this table is reliable at between the .05 and .01 level of confidence.

The obtained difference in retention between experimental and control groups, although statistically significant, would undoubtedly have been even greater if the learning passage used for matching purposes had been in the same subject matter field as the steel material (i.e., if the relationship between the two sets of scores were higher than that indicated by the correlation of .64 between the steel and pubescence scores). Another experimental condition probably detracting from the difference between the two groups was the fact that the steel material was not completely unfamiliar to many Ss. Because of some prior general familiarity with the contents of the steel

[2] Appreciation is expressed to Pearl Ausubel for assistance in the processing of the data.

[3] The standard error of the difference for equated groups was calculated according to a method described by Edwards (1954, pp. 282–288).

passage, many *Ss* already possessed relevant and stable subsuming concepts. These obviously rendered less significant the potential learning advantages conferrable by advance organizers.

It could be argued, of course, that exposure to the experimental introduction constituted in effect a partial substantive equivalent of an additional learning trial. Actually, however, any substantive repetition was at most very indirect, since the introductory passage consisted of much more inclusive and general background material than was contained in the learning task itself, and also provided no direct advantage in answering the test items. Furthermore, according to behavioristic (interference) theory, prior exposure to similar but not identical learning material induces proactive inhibition rather than facilitation.

Advance organizers probably facilitate the incorporability and longevity of meaningful verbal material in two different ways. First, they explicitly draw upon and mobilize whatever relevant subsuming concepts are already established in the learner's cognitive structure and make them part of the subsuming entity. Thus, not only is the new material rendered more familiar and meaningful, but the most relevant ideational antecedents are also selected and utilized in integrated fashion. Second, advance organizers at an appropriate level of inclusiveness provide optimal anchorage. This promotes both initial corporation and later resistance to obliterative subsumption.

The appropriate level of inclusiveness may be defined as that level which is as proximate as possible to the degree of conceptualization of the learning task—relative, of course, to the existing degree of differentiation of the subject as a whole in the learner's cognitive background. Thus, the more unfamiliar the learning material (i.e., the more undifferentiated the learner's background of relevant concepts), the more inclusive or highly generalized the subsumers must be in order to be proximate. If appropriately relevant and proximate subsuming concepts are not available, the learner tends to use the most proximate and relevant ones that are. But since it is highly improbable, however, that we can count on the spontaneous availability of the most relevant and proximate subsuming concepts, the most dependable way of facilitating retention is to introduce the appropriate subsumers and make them part of cognitive structure prior to the actual presentation of the learning task. The introduced subsumers thus become advance organizers or anchoring foci for the reception of new material.

Even though this principle seems rather self-evident it is rarely followed in actual teaching procedures or in the organization of most textbooks. The more typical practice is to segregate topically homogeneous materials into separate chapters, and to present them throughout at a uniform level of conceptualization in accordance with a logical outline of subject matter organization. This practice, of course, although logically sound is psychologically incongruous with the postulated process whereby meaningful learning occurs, i.e., with the hierarchical organization of cognitive structure in terms of progressive gradations of inclusiveness, and with the mechanism of accretion through a process of progressive differentiation of an undifferentiated field. Thus, in most instances, students are required to learn the details of new and unfamiliar disciplines before they have acquired an adequate body of relevant subsumers at an appropriate level of inclusiveness.

As a result, both students and teachers are often coerced into treating meaningful materials as if they were rote in character, and students consequently experience unnecessary difficulty and reduced success in both learning and retention. The teaching of mathematics and science, for example, still relies heavily on rote learning of formulas and procedural steps, on recognition of stereotyped "type problems," and on mechanical manipulation of symbols. In the absence of clear and stable concepts which can serve as anchoring points and organizing foci for the incorporation of new meaningful material, students are trapped in a morass of confusion and have little choice but to rotely memorize learning tasks for examination purposes. The traditional historical introduction to new and primarily nonhistorical, subject matter concepts possibly enhances student interest, but lacks the necessary substantive content to serve this organizing function (see examples under Procedure section above).

The pedagogic value of advance organizers obviously depends in part upon how well organized the learning material itself is. If it contains built-in organizers and proceeds from regions of lesser to greater differentiation (higher to lower inclusiveness), rather than in the manner of the typical textbook on lecture presentation, much of the potential benefit derivable from advance organizers will not be actualized. Regardless of how well-organized learning material is, however, it is hypothesized that learning and retention can still be facilitated by the use of advance organizers at an appropriate level of inclusiveness. Such organizers are available from the very beginning of the learning task, and their integrative properties are also much more salient than when introduced concurrently with the learning material.

## Summary and Conclusions

An empirical test was made of the hypothesis that the learning and retention of unfamiliar but meaningful verbal material could be facilitated by the advance introduction of relevant subsuming concepts (organizers). Experimental and control groups of 40 undergraduate Ss were equated on the basis of sex, field of specialization, and ability to learn unfamiliar scientific material. The learning task consisted of a 2,500-word passage of empirically demonstrated unfamiliarity, dealing with the metallurgical properties of steel. On two separate occasions, 48 hours and immediately prior to contact with the learning task, experimental Ss studied a 500-word introductory passage containing substantive background material of a conceptual nature presented at a much higher level of generality, abstraction, and inclusiveness than the steel material itself. This passage was empirically shown to contain no information that could be directly helpful in answering the test items on the steel passage. Control Ss similarily studied a traditional type of historical introduction of identical length. Retention of the learning material was tested 3 days later by means of a multiple-choice test. Comparison of the mean retention scores of the experimental and control groups unequivocably supported the hypothesis.

The facilitating influence of advance organizers on the incorporability and longevity of meaningful learning material was attributed to two factors: (*a*) the selective mobilization of the most relevant existing concepts in the learner's cognitive structure for integrative use as part of the subsuming focus for the new learning task, thereby increasing the task's familiarity and meaningfulness; and (*b*) the provision of optimal anchorage for the learning material in the form of relevant and appropriate subsuming concepts at a proximate level of inclusiveness.

The suggestion was offered that the greater use of appropriate (substantive rather than historical) advance organizers in the teaching of meaningful verbal material could lead to more effective retention. This procedure would also render unnecessary much of the rote memorization to which students resort because they are required to learn the details of a discipline before having available a sufficient number of key subsuming concepts.

## REFERENCES

Ausubel, D. P., Robbins, Lillian, C., & Blake, E., Jr. Retroactive inhibition and facilitation in the learning of school materials. *J. educ. Psychol.*, 1957, 48, 334–343.

Edwards, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.

# 35

# Retention and Transfer in Children of Low, Average, and High Intelligence

## — *Herbert J. Klausmeier and John Check*

*Calling for experimentation with human subjects in school settings in order to increase our knowledge of classroom learning phenomena, the authors report an investigation of retention and transfer abilities among school children of high, average, and low ability. A socially useful problem (an arithmetic one), graded in difficulty according to the children's present achievement levels, was employed as the task to be learned. Results indicated no significant differences among the three ability groups, either in retention or in transfer. Implications for research design and classroom practice are discussed.*

THE LACK of comprehensive knowledge about learning phenomena in classroom settings requires experimentation with human subjects in school settings. Further, since education is concerned with socially useful learnings, the results of laboratory experiments which employ nonsense rather than meaningful learning materials must be reexamined for their applicability to classroom situations. In this study, retention and transfer abilities were compared among children of low, average, and high intelligence. Each child learned to solve a socially useful problem, graded in difficulty to his present achievement level; and subsequent measures of retention and transfer were secured and analyzed.

Reprinted from *Journal of Educational Research*, 1962, 55, 319–322, by permission of the senior author and the publisher.

## Procedure

### Subjects

The subjects were 20 boys and 20 girls of low intelligence (WISC IQ 50–80), 20 boys and 20 girls of average intelligence (WISC IQ 90–110), and 20 boys and 20 girls of high intelligence (WISC IQ 120–146). The low IQ children were drawn from fourteen special classrooms for mentally retarded educable children. Any low IQ child who showed characteristics of organic involvements such as monogolism, or who was severely handicapped orthopedically or in hearing, vision, or speech was exluded from the study. Similarly, no average or high IQ child exhibited any serious physical defect. The average and high IQ children were drawn from fifteen regular fifth-grade classrooms.

All children had birthdates between September 15, 1947, and December 15, 1948. The mean age of children at the midpoint of this study was 131 months.

### Method

A group of twenty-nine problems at three main levels of difficulty was arranged, one difficulty level for each IQ group. After extensive experimenting with pupils in four fifth-grade classrooms and with a small group of low IQ children, none of whom were subjects of the study, the problems for low IQ children were further subdivided with the lowest level intended for children who did not know the number of cents in a dime or who could not do simple addition. A reserve level of problems of high difficulty was established for any high IQ child who might solve the assigned problem in less than three minutes. The assignment of problems within each difficulty level to individual children was made randomly, in order to carry out a combined retention and transfer design, described later.

Table 1 shows the number of children whose final problem was at each level, and it gives a sample problem at each difficulty level. Eleven low IQ

TABLE 1. Number of Children Whose Problem Was at Each Level of Difficulty

| Level of Difficulty | Amount Required | No. of Coins and/or Bills to Be Used | Low IQ Boys | Low IQ Girls | Average IQ Boys | Average IQ Girls | High IQ Boys | High IQ Girls |
|---|---|---|---|---|---|---|---|---|
| A–1 | 8¢ | 4 | 11 | 13 | | | | |
| A–2 | 13¢ | 4 | 2 | 5 | | | | |
| A–3 | 47¢ | 5 | 7 | 2 | | | | |
| B | 57¢ | 9 | | | 18 | 19 | | |
| C | $1.97 | 9 | | | 2 | 1 | 19 | 19 |
| R[*] | $14.53 | 16 | | | | | 1 | 1 |

[*] Level R was prepared for high IQ children who solved their originally assigned problem at level C in less than three minutes.

boys and thirteen low IQ girls performed at the lowest level of difficulty, and one of three problems at this level was to make eight cents with four coins. One high IQ boy and one high IQ girl were at the R, reserve level of difficulty. All problems in Level A required making a specified amount of money with the least possible number of coins; all problems in other levels required using a larger number of coins; e.g., fifty-seven cents with nine coins. The average and high IQ children used only paper and pencil in solving the problem; however, actual coins were used by the low IQ children.

In the teaching-learning situation, the examiner demonstrated how to make a specified amount of money with a specified number of coins, such as six cents with two coins. Following the demonstration, the examiner presented the child his assigned problem and started the stop watch. If the child solved the problem in less than three minutes, he was given a problem at the next higher level and time was started over. Cues or helps were provided at the end of six, nine, and twelve minutes according to a prearranged schedule in order to make certain that each child reached a solution to the problem in fifteen minutes. By design, the possible range of time to complete the problem was between three and fifteen minutes.

Immediately after the solution of the problem, a five-minute period of controlled interpolated activity started during which the child listened to a recorded story and several songs. Then the original problems were presented again to 60 of the children (retention group) and the time to solve and the cue level were noted. Seven weeks later, the same task with identical directions was presented to the 60 children in the retention group. Now follows a description of how a transfer group of 60 children were treated.

The 120 children were randomly divided into two groups, one designated a retention group and the other a transfer group. Thirty boys and thirty girls, ten from each IQ level, were randomly drawn and assigned transfer problems. Each of these 30 boys and 30 girls was assigned a different problem but at the same level of difficulty as he had originally learned. For the five-minute transfer measure then, the 60 children designated as transfer subjects received different problems, whereas the 60 children designated as the retention subjects were given the identical problems they had learned originally. Similarly, for the seven-week retention measure the retention subjects solved the same problems as before; whereas the transfer subjects were assigned new problems.

## Results

The design of this study required the problems to be graded to each child's achievement level. The means and standard deviations of the acquisition, five-minute, and seven-week retention and transfer scores in seconds are

TABLE 2. Means and Standard Deviations of the Time-to-Learn, Five-Minute and Seven-Week Retention and Transfer Scores

| | | Time-to-Learn in Seconds | | 5-Minute Retention in Seconds | | 7-Week Retention in Seconds | |
|---|---|---|---|---|---|---|---|
| | N | Mean | S.D. | Mean | S.D. | Mean | S.D. |
| Retention | | | | | | | |
| Low | 20 | 569.20 | 256.77 | 107.55 | 214.14 | 189.00 | 246.71 |
| Average | 20 | 590.30 | 252.31 | 84.70 | 183.16 | 362.75 | 298.42 |
| High | 20 | 541.70 | 238.44 | 145.85 | 241.67 | 233.75 | 212.47 |
| Transfer | | | | | | | |
| Low | 20 | 679.15 | 146.73 | 501.25 | 303.01 | 415.90 | 287.72 |
| Average | 20 | 665.00 | 252.35 | 457.95 | 292.26 | 467.35 | 293.00 |
| High | 20 | 615.25 | 246.83 | 494.95 | 310.34 | 378.40 | 275.92 |
| Method | | | | | | | |
| Retention | 60 | 567.06 | 249.29 | 112.69 | 214.32 | 261.83 | 254.98 |
| Transfer | 60 | 653.13 | 220.70 | 484.71 | 301.95 | 420.55 | 285.64 |

given in Table 2. The mean time-to-learn for the low, average, and high IQ retention groups was much alike—569, 590, and 542 seconds, respectively; the difference among the IQ groups was not significant at the .05 level.

The results of the four analyses of covariance (Edwards, 1950) of the acquisition and the retention or transfer measures with acquisition the covariate are shown in Table 3. The largest F of 2.13 does not approach significance at the .05 level. Thus, whether the relationship is between acquisition and relearning or between acquisition and transfer, the difference among the three IQ groups is not statistically significant at the .05 level. This

TABLE 3. Analysis of Covariance for Acquisition, Five-Minute, and Seven-Week Retention and Transfer

| Analysis Area | Source of Variation | Sum of Squares of Errors of Estimate | df | Mean Square | F | Significance of F |
|---|---|---|---|---|---|---|
| Acquisition 5-Min. Ret. | Adjusted means | 59326.3 | 2 | 29913.1 | .67 | $P > .05$ |
| | Within groups | 2486921.3 | 56 | 44409.3 | | |
| | Total | 2546747.6 | 58 | | | |
| Acquisition 7-Wk. Ret. | Adjusted means | 286115.8 | 2 | 143057.9 | 2.13 | $P > .05$ |
| | Within groups | 3747802.6 | 56 | 66925.0 | | |
| | Total | 4033918.4 | 58 | | | |
| Acquisition 5-Min. Transfer | Adjusted means | 26032.1 | 2 | 13016.0 | .12 | $P > .05$ |
| | Within groups | 5936295.8 | 56 | 106005.2 | | |
| | Total | 5962327.9 | 58 | | | |
| Acquisition 7-Wk. Transfer | Adjusted means | 78070.0 | 2 | 39035.0 | .42 | $P > .05$ |
| | Within groups | 5173764.0 | 56 | 92388.6 | | |
| | Total | 5251834.0 | 58 | | | |

ever, Hendrix (5) found that with a mathematical principle those groups that discovered the principle independently and left it unverbalized exceeded those who discovered and then verbalized, and both exceeded in transfer those who had the principle stated for them and then illustrated.

Opposed to Katona and Hendrix are those like Craig who concluded: "The more guidance a learner receives, the more efficient his discovery will be; the more efficient discovery is, the more learning and transfer will occur" (*1*, p. 72). In a further study with college groups and with the same method of having the S pick out that alternative among five which does not fit a principle (2) he verified that significantly more such problems were solved when the principle was stated above it than when the S was given only the instruction that one of the five items did not belong. One should note, however, that he found no difference between his groups on transfer to new principles nor was there any difference in retention after 3 or 17 days, although at 31 days the difference favored the directed group.

While Craig's experimental results actually gave little or no support to his claim that guidance is desirable for transfer, more serious opposition to Hendrix and Katona came from Kittell (*8*). He found that "intermediate direction" (starting a principle) was significantly superior to both the "minimal direction" (told only that one of five alternatives would not fit) and "maximal direction" (*E* told the principle and worked out the answer for S), with minimal direction definitely the inferior method. His subjects were sixth graders while Craig's were college students. That difference in age and educational level may explain the contrast in results. Also Kittell's low number of successful solutions (means only 4.59 for intermediate direction and 1.93 for minimal direction out of 15 principles) suggests that the problems based on linguistic arrangements and meanings may have been too difficult for the Ss. If that were the case, then following directions in the stated principle was about the only way to solve the problem when unprovided with sufficient apperceptive mass and experience. Haslerud (*4*) found that while naive rats transferred anticipatively from forced turns near the goal into prior free units of a maze just as well as when those goal turnings had been established by trial and error, only active trial and error cul-de-sac elimination in the goal region could readjust an established pattern in the prior free units. If a similar limitation on effectiveness of guidance is present in human Ss, then one might expect any advantage primarily in young Ss and that mainly on their initial learning but none for memory and transfer where Ss have sufficient background to derive a solution themselves.

While the Katona and Hendrix concept of how to get maximal transfer seems to have face validity, at least for adults, their controls and statistical supports are unsatisfactory. When one draws his conclusions on the basis of one principle, e.g., the sum of the first *n* numbers, a question remains of how much of the conclusion is a function of the particular problem used or the selection of individuals for the various groups. More convincing differentiation of principle given from principle derived would seem to require homogeneously varied problems posed in quantity to the same individuals. A likely material has been found in an extension of the familiar cryptogram "Come to London" in the Stanford-Binet. An unpublished pilot study by the junior author under the senior author's supervision indicated an advantage for memory of the independent solving of such coding principles. The present study extends a similar method to transfer. The hypothesis tested is that principles derived by the learner solely from concrete instances will be more readily used in a new situation than those given to him in the form of a statement of principle and an instance.

## Procedure

Subjects for the experimental group were 76 members, ranging from freshmen to seniors, of two general psychology classes at the University of New Hampshire. The control group of 24 students in another psychology class ranged from sophomores to seniors.

The experimental groups were each given two coding tests, the second being administered one week after the first. The control group was given only the second test. All tests were administered by the senior author.

The first test composed of 20 coding problems was designed to give the students two types of experience: (*a*) problem solving with specific directions for deciphering the code printed above each problem, and (*b*) problem solving with no directions given. The first part of each problem was the four-word sentence "They need more time," followed by the same sentence in code. A different code was used in each problem. The second part of each problem was the four-word sentence "Give them five more," which the Ss were asked to translate into the code for that problem. The given and derived problems were alternated so that the S would solve approximately equal numbers of each kind. As a control for differences between the codes, there were two test forms, A and B. The same codes were used in both, but those for which directions were given in form A had to be deciphered by the S in form B, and vice-versa. The problems were arranged in approximately the apparent order of difficulty. Examples of moderately easy coding rules are: "For each letter of the sentence write the letter that follows it in the alphabet." "Write the first two letters of each word and then the last two letters of each word." To introduce the test the senior author told the Ss that the test was an experiment in cryptography. He wrote an illustrative code on the blackboard and purposely worked it out partly incorrectly to encourage remonstrances from the group that a system or principle was possible. Ss were asked to solve the problems in the order they appeared on the test and to do as many as they could in the time allotted. Since the 45 minutes allotted was ample time for all but one or two students in each group, the test was essentially a power rather than a speed test. The Ss were not told that they would be retested on the same material.

The second test printed only in one form was given to both the experimental and control groups. Again, the 20 codes used in the first test were used, but instead of the common sentence of Test 1, there were 20 different English sentences 14 to 18 letters in length followed by four translations into code. Only one translation was correct, and the Ss were asked to check it. They were told that the other three were simply letters arranged in random order. They were not told that numbers had been assigned to letters of the alphabet and that letters for two of the four codes had been selected according to the order in which those numbers appeared in a list of random numbers. The third false code was composed of letters of the English sentence arranged according to random numbers. The order in which the four codes followed the sentence and the order in which the problems were arranged on the test were also random. No mention was made of the previous test, nor was the purpose of the test told until both tests had been given and the results compiled.

## Results

The data for each individual in the experimental group consisted of four scores: (*a*) Number of correct codings on Test 1 problems where the rule was given, hereafter called $G_1$ scores. (*b*) Number of correct codings on Test 1 problems where the coding principle had to be derived by the S, hereafter called $D_1$ scores. (*c*) Correct alternatives for those codes in Test 2 that had been G type in Test 1. (*d*) Correct alternatives for those codes in Test 2 that had been D type in Test 1. In the control group the score was the total number of correct alternatives on Test 2. Any coding was considered correct if no more than 1 of the 16 letters was wrong, since carelessness rather than lack of understanding of the principle was probably responsible for the lone error.

Since there was no difference between their results, the two experimental classes were combined. The analysis of results, however, was carried through separately for Forms A and B of Test 1 because a difference significant at the .05 level indicated that the 10 odd and the 10 even problems had not been exactly equated for difficulty. Nevertheless, the direction of results for both A and B groups showed equally high differentiation of G and D situations.

Test 2 performance of the experimental group was significantly different from that of the control group. The means, 15.74 and 10.75 respectively, differ beyond the .001 level. Apparently something is transferred from the Test 1 experience.

The crucial comparisons are between the G and D kinds of problems. For both Forms A and B on Test 1, significantly more G problems were correctly coded: 8.86 and 8.36 against 5.86 and 4.88 for G and D respectively. The results for Test 2 a week later are given in Table 1. If the differences are added algebraically to the Test 1 scores given in the previous sentence, one obtains the nearly equal transfer scores of Craig's experiment (2). But since each individual was his own control for both G and D problems on Tests 1 and 2, it is legitimate to use the subtraction method to find the standard error of the difference for paired observations. The correct identification of those codes which had been D type on Test 1 increased 46% while those which had been G decreased 10%. Both changes are significant, at the .001 and .05 to .01 levels respectively. There is reason to think that both curtailing time to make Test 2 a speed test and increasing time to greater than a week between the learning of the codes on Test 1 and the transfer on Test 2 would accentuate the differences.

TABLE 1. Difference in Number of Problems Successfully Coded Between the Transfer Test (Test 2) and the Initial Learning (Test 1) with Each Individual as His Own Control

| | $N$ | $\bar{x}$diff | $\partial$diff | $\partial\bar{x}$diff | $t$ | signif. |
|---|---|---|---|---|---|---|
| $D_2$–$D_1$ Problems | | | | | | |
| Test A | 36 | 2.83 | 3.34 | .56 | 5.06 | $p < .001$ |
| Test B | 40 | 2.50 | 3.30 | .52 | 4.81 | $p < .001$ |
| $G_2$–$G_1$ Problems | | | | | | |
| Test A | 36 | −0.83 | 2.37 | .38 | 2.17 | $p < .05$ |
| Test B | 40 | −1.03 | 2.21 | .35 | 2.94 | $p < .01$ |

## Discussion

This experiment has added strong support to the contention of Katona and Hendrix that independently derived principles are more transferable than those where the principle is given to the student. Even though Ss produced more correct codings on the original learning when the principle was stated for them, on the "payoff," or "applying" to use Katona's term, the advantage definitely passed to those principles derived by the student himself. Fast and accurate learning or performance under immediate guidance is no guarantee of transfer to new problems without such support. From

Craig's and our experiments the conclusions just stated are supported by results on college students, but testing of grammar level students by principles of a more suitable level of difficulty than used by Kittell (8) might show a wider application. Our coding method could be easily adapted for that purpose.

The obtained results of this experiment do not follow from inadequate controls. The alternate Forms A and B allowed each principle to be given (G) and derived (D). Individual differences with respect to problem solvinb in the Ss were ruled out since each person responded to 10 G and 10 D problems on Test 1 and the follow-up of each of these on the transfer Test 2. The control group's much poorer performance on Test 2 indicated that a genuine transfer function was present. Making time on each test practically unlimited pushed the G and D types of presentation to their limit as power tests.

Two possible weaknesses in the transfer Test 2 need to be examined. With four alternatives for each problem, a chance score would average 5. The control group had 10.75 problems correct; this showed good adaptation to the test but significantly less than the 15.74 of the experimental groups. The question whether the better performance of the experimental group was just the result of a second session of practice on coding problems can probably be answered by reference to the study by Warren (9). He found that adults on letter-symbol substitution rapidly attain a plateau on transfer problems because of "learning sets" from early childhood. Coding is in that class of simple activities for adults where experience and practice as such make little difference after the first 10 minutes. Even if one took the maximum change of 37% during Warren's 16 five-minute periods, it would be less than the nearly 50% advantage of our experimental group over the control group. The second possible weakness arises from the randomized construction of the false alternatives of the transfer test. It is conceded that a person might try to solve the problems by excluding the three alternatives because of their random characteristics rather than by trying to recognize and verify some consistent principle in the one true alternative. However, the principles must have played a significant role in the solutions because without them the results of the control and experimental groups would have been equal since they had the same instructions and equal opportunity to use this abortive device.

The theories of transfer found in current educational psychologies are inadequate to explain the present experiment. The senior author plans to develop in another place a theory that transfer is fundamentally an anticipative rather than a perseverative function and that to get transfer one must always counteract the finality of a goal (3). A stated principle to some extent, and even more Kittell's "maximum guidance" of *E* doing the problems for S after giving him the principle, practically stops transfer, like other goals. Hendrix (*6*) states from Thorndike that only 5% of high school students have language ability sufficient to receive a ready-made sentence and find readily illustrations in their own background to provide the prerequisite to meaning. If the results of the present experiment can be verified for a wider range of ages and apperceptive masses, then the implications for a direct attempt to teach for transferable principles can not be neglected.

## Summary

The educationally important question of how much guidance is desirable if one is interested in transfer was tested experimentally by a new use of coding. Each of 76 college students as his own control translated into 20 different codes a common four-word sentence, with the rule given for half of the problems and required to be derived solely from example for the other half. As in previous studies on initial learning, the Ss did significantly better on those problems with the rule given. However, a week later on a multiple-choice transfer test consisting of 20 different sentences, one for each of the 20 coding principles of the first test, the selection of the adequate code from three specious ones made by randomizing letters gave very different results. The scores were significantly increased for those problems which had formerly been derived as contrasted with a significant decrease for those problems where the rule had formerly been given. A control group of 24 college students given only the second test proved by significantly poorer performance than the experimental group the value of transfer from the first test. The results give strong support to the postulate of Hendrix that independently derived principles are more transferable than those given. The apparent contradiction with Kittell's study of children was explained by the smaller apperceptive mass in the child, and the prediction was hazarded that as naivety is lost, the probability of transfer from learning which is minimally directed is increased.

## REFERENCES

1. Craig, R. C. *The transfer value of guided learning.* New York: Teachers College, Columbia Univer. Bureau of Publications, 1953.
2. Craig, R. C. Directed versus independent discovery of established relations. *J. educ. Psychol.*, 1956, **47**, 223–234.
3. Haslerud, G. M. Properties of bi-directional gradients at subgoals. *J. genet. Psychol.*, 1950, **29**, 67–76.

4. Haslerud, G. M. Anticipative transfer of mechanically guided turns. *J. exp. Psychol.*, 1953, 45, 431–436.
5. Hendrix, Gertrude. A new clue to transfer of training. *Elem. Sch. J.*, 1947, 48, 197–208.
6. Hendrix, Gertrude. Prerequisite to meaning. *Math. Teacher*, 1950, 43, 334–339.
7. Katona, G. *Organizing and memorizing*. New York: Columbia Univer. Press, 1940.
8. Kittell, J. E. An experimental study of the effect of external direction during learning on transfer and retention of principles. *J. educ. Psychol.*, 1957, 48, 391–405.
9. Warren, J. M. Intertask transfer in code substitution learning. *J. genet. Psychol.*, 1956, 89, 65–70.

# 37

# Transfer Effects of Learning a Verbal Generalization — *Susan M. Ervin*

*Using third- and fourth-grade children as subjects, the author reports an investigation of the effects on transfer of learning a verbal generalization. The study, it is concluded, confirms the theory that verbal training facilitates transfer provided several necessary conditions are present. The author suggests that children must be able to see the similarities between two situations and must be able to link generalizations to action.*

IN EDUCATIONAL practice, curriculum planning assumes that there will be a saving in the learning of related material. But the content of the learning to be transferred is seldom considered specifically enough for precision in planning. Further, most laboratory experiments use shorter time intervals between the training and transfer situations, and more similarity between the situations, than is the usual case in natural situations.

The term transfer has been applied to any learning experiment in which training and test conditions are not identical (*11, 17*) and where there has been concern with generalization, rather than with manipulating strength or frequency alone. This indicates one reason for slow progress in such research; it has been concerned with the dimension of similarity, which is more difficult to quantify than strength or frequency of response. In most applied research, complex stimuli and responses are used, which present maximum difficulties in advance specification of similarity quantitatively.

Reprinted from *Child Development*, 1960, 31, 537–554, by permission of the author and the Society for Research in Child Development, Inc.

These hurdles appear in studies in which the common feature in training and test is complex visual-motor coordination (3), verbal training on nonsense syllables (13, 16, 18), or stimulus pre-differentiation (1, 4, 5, 6, 7, 14).

The study of verbal or motor mediation (9) has overtly certain similarities to transfer research. Characteristically, responses come to be elicited by new stimuli just as in transfer research. But the new stimuli are not similar to the old; they are linked to old responses by training. In the classic mediation models one can specify quantitatively the probabilities that each link in the mediation series will elicit the next because training has been controlled. One assumes that overt response probabilities also hold for implicit mediating responses. For certain types of materials this assumption is clearly fallacious—e.g., taboo words (8, 10); but the assumption applied to other verbal material has been reasonably well supported (2, 12, 15).

Design in mediation studies may follow one of a variety of patterns. Only two will be discussed below. In these descriptions S represents some stimulus—a light, a complex visual pattern, a color, a sound. R represents a response—a word to be said, pushing a lever. V represents a stimulus which can be presented either by another person or by the subject as a self-stimulating response. Verbal responses are ideal in this role. The two patterns follow:

**1.**

$S \rightarrow V$

$V \rightarrow R$

Test:

$S \rightarrow (V) \rightarrow R$

**2.**

$V \rightarrow R$

$$\begin{tikzcd} S \arrow[r] \arrow[dr] & V \arrow[d] \\ & (R) \end{tikzcd}$$

Test:

$$\begin{tikzcd} S \arrow[r] \arrow[dr] & (V) \arrow[d] \\ & R \end{tikzcd}$$

The parenthetical responses are not overt. V, it will be noted, functions both as stimulus and as response. When it occurs as a response, it has, like all responses, self-stimulating properties. The increase in S-R occurrences after training is attributed to implicit V responses during the test. In the second paradigm, training could also occur due to implicit R responses during learning. The second paradigm is often used for word associative material, in which the training in V → R is prior to the experiment and is measured by response frequencies in the word association test, and both V and R are verbal.

The common feature of all mediation experiments is that transfer will not occur unless the implicit response occurs. One must know its probability in order to predict the result.

This point is very important in applied situations, where verbal mediation is often used to produce transfer. Suppose, for example, the observable outcome of training in a certain school subject is not verbal, but consists in drawing, action, or symbol manipulation as in the case of natural sciences or mathematics. Frequently, verbal mediation is assumed to be present, and teaching is aimed at altering interverbal probabilities. To simplify in the extreme, let us represent some stimulus pattern as S, the name elicited by it as N, a proposition about it or verbal response as P, and an action elicited by this proposition or verbal response as R. For instance: S, blinking light on traffic signal; N, "blinking light"; P, "slow down at a blinking light"; R, slowing down.

Verbal training is characterized by manipulating the relationship of N and P. Children are taught verbal rules, descriptions of operations, reading about science. Drivers are given verbal instructions or booklets containing traffic regulations.

Under such a regime transfer to the S-R situation will occur only under the following conditions:

1. The stimulus actually elicits verbal mediating responses. The driver who has never before encountered blinking lights and has only been told to stop at them will not stop unless blinking lights elicit a verbal mediator. In children, with less verbal training than adults, the probability of verbal mediation may generally be less.

2. The implications for action are clear. That is, there is a strong likelihood that the verbal formulation here called P will elicit response R. In the case of a verbal formula "slow down" the probability of slowing a car down is high in a person who knows how to drive. But frequently rules characterize a situation without specifying appropriate responses for the actor. Then the uncertainty of his response is increased by uncertainty in any of the factors that enter into the link. One could test the probability of R, given P, by giving a person overtly the verbal formulation believed to occur mediately.

3. The skills required in performing R are fully available. Where P refers to a perceptual-motor act, improved performance of R rests both on increasing the probability of P and on the skill of the subject in performing R.

The above description is a fragmented one which suggests a sequence. While analytically a sequential description is the most useful, it is quite likely that these processes occur with near-simultaneity if they are well learned. To argue from a phenomenological standpoint that we are not aware of implicit language is an inadequate objection, just as an observation that we are unaware of the kinesthetic sensations in typewriting or piano-playing. Probably these near-simultaneous processes occur so fast they are experienced as insight, and once they have occurred in a given set of circumstances successfully the probability of their reoccurrence and of a short-circuiting of the mediating process will be higher.

In the following experiment two variables are distinguished: trained verbal mediation, in contrast to mediation training without a trained verbal response; the other is the difference in response probabilities of verbal responses and nonverbal responses in the test situation—that is, the predicted increase in R after training is made both for motor performance and verbal. Both the verbal and the nonverbal training situations in addition relied on generalization since the training and transfer tasks were similar but not identical. The following hypotheses are tested:

1. There will be an increment to performance resulting from trained verbal mediation, beyond that gained from relevant nonverbal training only. This training consists both of label training ($S \rightarrow N$) and of interverbal "rule" training ($N \rightarrow P$).

2. There will be a greater effect on verbal than on nonverbal performance as a result of trained verbal mediation. This is particularly likely to be true when the implications of the rule for action are not specified and there has been no special training to guarantee that the verbalization will generate changed nonverbal responses with high probability.

3. Those who show improved verbal performance will more often show better motor performance. If verbal responses be considered evidence that verbal mediation must have occurred, at least the necessary, but not the sufficient conditions for improved motor performance may be present.

4. Among those not given explicit verbal training, spontaneous verbal comments like those offered in the training will lead to an increment in nonverbal and verbal performance in the transfer task.

## Method

### Subjects

The subjects were 48 third- and fourth-grade children ranging in age from 8–1 to 10–1. The mean age of the girls was 9–3 and of the boys was 9–1.5. The groups were originally selected on the basis of a Gates Reading Test IQ so that half of the children were above 120 in IQ and half below in each subgroup. The subjects were divided into three groups matched on age, sex, grade, and IQ, which were the three experimental groups.

Subsequently, after the experiment was completed, it was decided that we wished to differentiate mental abilities, and three subtests of the Primary Mental Abilities Test were administered. It then appeared that the nonverbal group in the third grade was significantly inferior in form reasoning to the other two groups and that in the fourth grade the control group was superior to the other two groups on form reasoning. In the third grade the verbal group was superior to the nonverbal in verbal ability, and in the fourth grade the control group was superior to the verbal group in verbal ability.

**Equipment**

The training apparatus consisted of a large plywood board surrounded on four sides by strips of molding and marked with a grid pattern. On one of the long sides a 2-inch thick plank 2 feet in length, with a rubber padding on its vertical surface, was rotated on a bolt at one end and set so that its vertical edge was 4 inches from the edge of the plywood. This was called the backboard. It had slotted arcs protruding at the back with wing bolts, so that it could be swung out at various angles from the back molding and clamped in place. Its full rotation was about 45°.

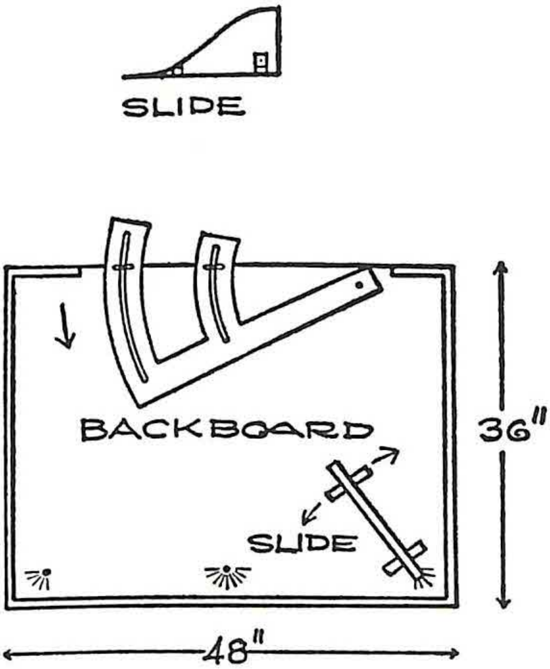


Fig. 1. *Training apparatus.*

On the opposite, long, side of the plywood surface was inserted a small slide like a miniature child's slide, with a maximum height of 6 inches. This slide was inserted on a pivot into holes placed in the board at the center and at each corner. In the corners it could be rotated 90° and in the middle, 180°.

A target ball was placed somewhere on the plywood surface in a position indicated by the grid markings. A small wooden ball could be rolled down the slide and, if the slide was aimed toward the backboard, the ball would ricochet.

The test apparatus consisted of a mirror in a position analogous to the backboard, but the whole structure was placed on a vertical dimension so that the "ground" or plywood board was vertical and protruding from it horizontally was a mirror. This was set in a cradle that could be rotated in an axis perpendicular to the board like the training backboard, and in addition the pivots on the mirror permitted rotation in an axis parallel to the board. A flashlight was set so that it was aimed up toward the mirror, but it could pivot out, away from or toward the board on an axis parallel to the

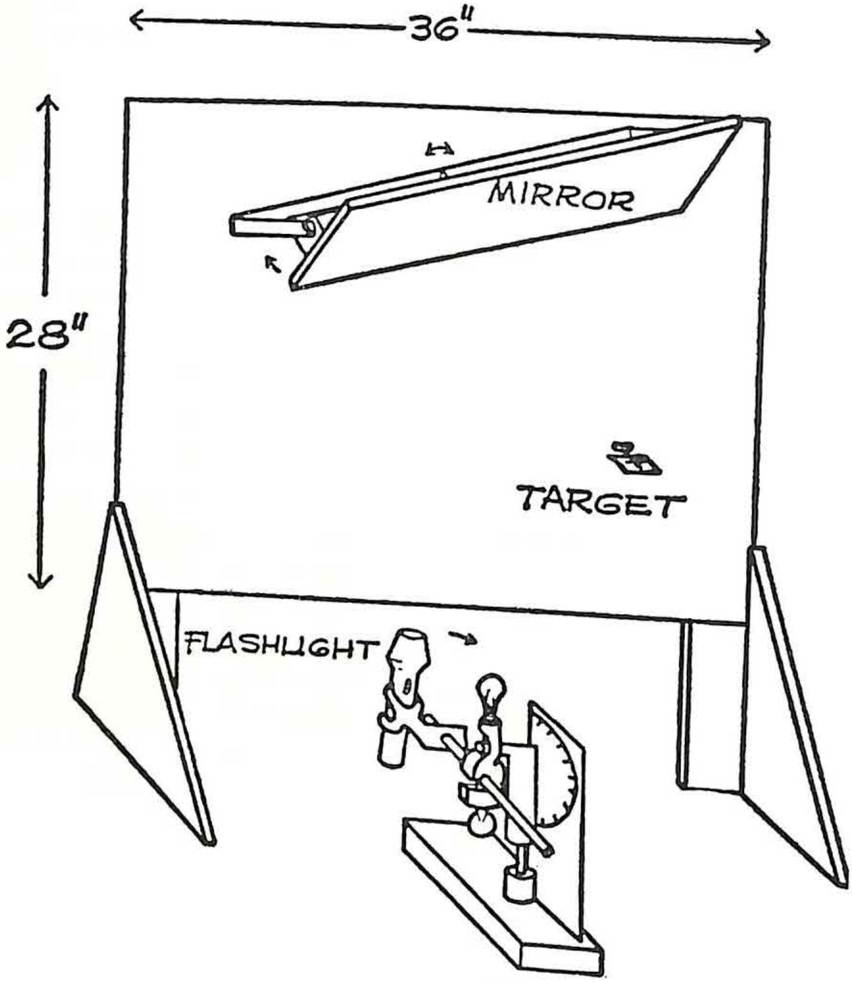


FIG. 2. *Test apparatus.*

board, or it could rotate on an axis perpendicular to the board as did the slide in the training apparatus. The board was covered with tin sheeting and a thin crinoline, and painted black. The target was a card with a cross, fixed to a magnet.

This target thus stood out on a horizontal plane from the vertical backboard. When the mirror was properly adjusted and the flashlight was aimed up toward it, the reflected light would fall on the red cross of reflector tape. When the flashlight was set directly under the board and pointed up, the path of light was visible along the uneven surface of the cloth up to the mirror and down again. This visible path was analogous to the visible path of the ball traced with flour during the slide procedure.

**Procedure**

The children were brought individually from the classroom to the experimental room in the school. For *training series A* they were shown the slide apparatus with backboard set parallel to the back molding and with the slide in center position. Instructions were given in the full 180° rotation of the slide and in proper handling for steadiness while the ball was being released. Four or five practice shots were allowed. The children were then told: "This is how the game works. I will put the target on the board. Then you have two tries to aim the slide so that the ball bounces back and hits the target. After two tries I will move the target and you get two more tries in the new place. I am going to keep a record of every time you hit the target and see how many prizes you get at the end."

Trials were continued to a criterion of five consecutive correct shots including no more than two near-misses, to allow for some error in the uneven bouncing of the ball. The target ball was placed in a series of positions on the board.

*Mediation training* was introduced, with two operations alike for the verbal and nonverbal training groups. First, the children were asked to draw the expected path of the ball on a piece of black construction paper placed in front of the backboard flat on the floor of the apparatus. During this procedure the slide was in the front left corner position, and the procedure was repeated three times with three different rebound angles. Each time after the path was drawn the ball was coated with flour and rolled down the slide. The flour traced a path on the paper.

After the first stage of mediation training the *nonverbal* group was told: "Now let's look at the pictures. I'm going to color them in, and then you look at them closely, so that you can learn to draw the way the ball goes." The experimenter then drew a line on each side of the point of contact, perpendicular to the backboard line, so that there were two congruent triangles formed by the angles of incidence and rebound.

The *verbal* group performed the same operation, but was told: "When a line turns like this it is called an angle. Here is the line of the backboard. Can you see three angles here? The backboard makes an angle here and here, and there is an angle in between. I am going to color in the side angles. Which angle looks more like this one? (Shows one of the six side angles in the three drawings) Like this one? (Shows another) So the two side angles made by the path the ball follows are always alike. They are equal. Whenever a ball bounces, the side angles it makes are alike."

In the second stage of mediation training the children were shown a series of mimeographed drawings of a horizontal line, representing the backboard, with a line projecting from it representing the oncoming ball. The children were asked where they thought the ball would go and drew the path. In the *nonverbal* group the child was asked: "Does your picture look like the pictures the ball made?" and, when he erred, was asked to fix the drawing so it would look like the picture the ball made. If he was wrong, the experimenter colored in the side triangles as before. The child continued to draw lines and correct them until he had drawn three consecutive correct ones. He was then told that if he remembered the pictures he would be able to do very well at the game.

The *verbal* group also was asked to draw pictures, but after each drawing he was asked if he had drawn the picture so that the two side angles were alike. At the end he was told: "When a ball bounces, it makes a path. Could you tell me what you learned about the two side angles it makes? The side angles made by the path are alike. If you think about that you will be able to do very well at this game."

*Training series B* required aiming of the slide from a position on the right of the board. The target ball occupied a fixed position but the backboard was pivoted into various positions. The criterion was the same as in series A. The nonverbal group was reminded to think about the pictures they drew, and the verbal group was asked to think about the rule about the path of the ball and was asked about it every five trials. For *training series C* the child set the backboard after the experimenter had moved the slide in rotation as in series B. At the end, each child was asked, "Could you tell me how you would describe the way you tried to make the ball go? Did you try to picture it to yourself at all? Did you imagine the path of the ball? Tell me if you can the rule that describes how the ball goes."

The *control* group never saw the slide apparatus. They were given training by the same experimenter on a completely unrelated task.

There were two test series with the mirror, approximately six weeks later. In *test series A* the child was asked to predict where a spot of light would appear when the flashlight pointed at the mirror was turned on. A series of 10 settings alternatively changed the tilt of the flashlight towards or away from the board, the tilt of the mirror to right or left, and the tilt of the mirror down or up towards the viewer in front of the apparatus. The spot might be cast as far as eight feet from the apparatus when the mirror and the flashlight both tilted out from the board. After the series, the child was asked: "If you had to explain to one of your friends how to tell where the spot was going to be, how would you explain? How did you usually decide where the spot would be? Did you picture anything, or did you have a rule about it?" The responses were tape recorded.

In *test series B* the target card was moved about the board, or the mirror was rotated to right or left, and the child was asked to set the flashlight so that the light when turned on would fall on the card. The child had two tries for each setting, with 14 settings. He was scored correct if he was right either time. In addition, the proportion of times he changed the flashlight in the correct direction after a wrong first setting was obtained from the second settings, after errors. After 10 settings on this series the child was asked for an explanation and shown the path of the light on the cloth.

A set of *mimeographed drawings,* showing the mirror, the flashlight, and the card, in various positions, was shown, and the child was asked to draw a line showing where the light would go and to tell whether the light would go on the card or not. Finally, all the children in the groups which had slide training were asked if the task reminded them of anything else and were asked what rule they learned about how to set the slide.

## Scoring

Performance scoring on the training series consisted of the number of trials to criterion, and on the test series consisted of the number of trials correct. The verbal statements were scored according to the following code:

*Correct complete rule.* The child states that the side angles are equal, or some equivalent of this statement, e.g., "The spaces on the sides are alike."

*Partial correct rule.* The child makes a statement about the way the light goes or a rule for judging a setting that excludes certain settings, and is correct as far as it goes, e.g., "It depends on how much more tilted the mirror is whether it is going to go higher or lower."

*Incorrect rule.* The child excludes certain possible settings with a rule that will sometimes produce wrong settings. "You aim at the corner of the mirror."

*Vague, incomplete statements.* The child makes a statement that is too vague to exclude any alternatives for the hearer, or says he doesn't know. "I try to imagine how the light will go." "The light goes up and then it comes back again."

The intercoder reliabilities of the above categories were satisfactory, ranging between 100 per cent agreement on the first category and 79 per cent on the second.

### Summary of Design

The design consisted of presentation of a training series and a test series to three groups of children. The test series was identical for all three groups. The control group was trained on an irrelevant task, while the verbal and nonverbal training groups received training with the slide apparatus and mediation training that differed only in that special terminology was introduced to the verbal group and in that they were taught a verbal rule.

## Results

### Effects of Training on Performance

No differences appeared between the two experimental groups and the control group in performance on the test transfer series. The startling performance of the fourth-grade boys in the control group led to an examination of various performance correlations with grade, sex, and spatial ability. On the *training series* boys and fourth graders did significantly better. In the nonverbal group all the PMA scores were positively correlated with slide task performance (Table 1). In the control group there was a significant tau of .48 between spatial ability and form reasoning scores and test (mirror) series performance.

TABLE 1. Correlation of Primary Mental Abilities and Performance (Tau)

| | Training Series | | |
|---|---|---|---|
| | A | B | C |
| Verbal Group ($N = 16$) | | | |
| Picture Vocabulary | −.39* | .06 | .06 |
| Spatial plus Form Reasoning | .20 | .11 | .03 |
| Nonverbal Group ($N = 16$) | | | |
| Picture Vocabulary | .34* | .66* | .18 |
| Spatial plus Form Reasoning | .13 | .53* | .28* |

* Tau significant beyond the .01 level.

These correlations suggest that boys, fourth graders, and children with higher spatial-form reasoning scores might have had outside encounters with the rebound phenomenon so that the special training had little effect. Rebound occurs, of course, in many kinds of ball games and in pool, which some of the children said they played. In addition, the control children had higher form reasoning scores than the others.

Therefore the girls, the third graders, and the children with low spatial and form reasoning ability were examined separately (Table 2) as a more untrained group than the others. In each case there was significant im-

TABLE 2. Median Performance Scores on Transfer Tests

| | | | | $p$° | |
|---|---|---|---|---|---|
| | Verbal | Nonverbal | Control | V vs. C | NV vs. C |
| Test A | | | | | |
| Total sample | 4 | 5 | 4 | *ns* | *ns* |
| 3rd Grade | 4 | 4.5 | 3.5 | *ns* | *ns* |
| Girls | 2.5 | 3 | 4 | *ns* | *ns* |
| Low Spatial and Form Reasoning | 4 | 4.5 | 4 | *ns* | *ns* |
| Test B | | | | | |
| Total sample | 9 | 9 | 7 | *ns* | *ns* |
| 3rd Grade | 9 | 9 | 6.5 | .08 | .01 |
| Girls | 8 | 9.5 | 7 | .097 | .014 |
| Low Spatial and Form Reasoning | 9 | 9 | 7 | $<.01$ | $<.001$ |

° Mann-Whitney one-tailed test.

provement in performance with nonverbal training, compared to the control group. Perhaps, the experimental training was an insufficient increment to the previous training of the other children.

The seventh setting on test series B was a key item. The mirror was tipped steeply, and the target card was low, near the flashlight. With the mirror thus tipped, the beam reflects up and to the left, quite high. The usual error is to aim the flashlight too high, thus sending the beam up to the ceiling.

TABLE 3. Performance on Key Item with Sharply Tipped Mirror

| | Treatment | | | | | |
|---|---|---|---|---|---|---|
| | Verbal | | Nonverbal | | Control | |
| | *C* | *U* | *C* | *U* | *C* | *U* |
| Immediate correct setting | 2 | 2 | 1 | 1 | 1 | 0 |
| Correct on second setting | 2 | 3 | 1 | 2 | 1 | 0 |
| Correct adjustment direction | 4 | 1 | 2 | 4 | 0 | 3 |
| Incorrect adjustment direction | 0 | 2 | 0 | 5 | 2 | 9 |

NOTE: Six-celled grouped $\chi^2$ of incorrect direction vs. others significant at .01 level.
*C* means a correct verbal formulation at the transfer session, series B.
*U* means an incorrect or unclear formulation.

This setting is important because it was an instance when an alternative setting rule usually followed would not work. Many children aim at a point somewhere between the vertical projection of the target and of the flashlight onto the mirror. When the mirror is relatively horizontal over the flashlight, this method is successful. When it is tipped steeply, it does not work. On a similar drawing only five of the 48 children drew correctly.

Thus, this setting is a sensitive test of whether the child is in fact balancing the angles of incidence and reflection and correctly visualizing the path of the light. The differences between groups in success on this single question are striking. Only two of the control group children were successful and few even adjusted the flashlight in the right direction for the second setting. Five of the children with nonverbal training and nine of those with verbal training were successful (Table 3). In these two groups none of the children who gave a correct verbal rule made an incorrect adjustment. The relation of verbal performance at the learning session and transfer to the key item may be seen in Table 4.

TABLE 4. Verbal Learning and Performance on Key Item with Sharply Tipped Mirror

| | Statements After Learning Session | | |
|---|---|---|---|
| | Verbal Formulation | No Verbal Formulation | Control |
| Immediate correct setting | 6 | 0 | 1 |
| Correct second setting | 6 | 2 | 1 |
| Correct adjustment direction | 8 | 3 | 3 |
| Incorrect adjustment direction | 2 | 5 | 11 |

NOTE: Grouped $\chi^2$ significant at the .05 level with $df = 2$.

## Effects of Training on Verbal Responses

The verbal training had a clearly facilitating effect on the series B verbal performances. The proportion of complete statements of the principle taught at the training session is greatest in the verbal group and least in the control group (Table 5). Incorrect formulations are most frequent in the control group and least frequent in the verbally trained group. Of the eight who did not offer a correct formulation in the verbal group, four could not remember what they had learned at the first session when asked. Verbal train-

TABLE 5. Verbal Statements at Transfer Session

| | Verbal | Nonverbal | Control |
|---|---|---|---|
| Series A* | | | |
| Complete rule | 2 | 0 | 1 |
| Partial correct | 4 | 8 | 5 |
| Incorrect | 1 | 0 | 1 |
| Other | 9 | 8 | 9 |
| Series B† | | | |
| Complete rule | 5 | 3 | 1 |
| Partial correct | 3 | 1 | 3 |
| Incorrect | 1 | 5 | 7 |
| Other | 7 | 7 | 5 |

* None significant.

† Verbal vs. Control for correct vs. incorrect yields $p$ of .025 by Fisher $t$ test. Verbal vs. Nonverbal for correct vs. incorrect yields $p$ of .064.

ing evidently does influence children's formulations or explanations in a transfer situation.

Some examples of the verbal formulations at the mirror session are these from the verbal group:

> *George* (7;1): I'd sorta put the flashlight so it goes at an angle. The other lady had a ball and you slide it off a thing and she had this other ball. . . . She told me it goes in an angle and it hits when it goes at a right angle. You put it at a right angle and it hits. . . . I took the thing and turned it around. If the ball was over here, the angles were the same so I'd go like this.
>
> *Ann* (7;9): It is going up now. It would hit the mirror and come down and the two sides would be even.
>
> *Ellen* (9;0): Well with the other game I guess it would be if they had a ping pong ball tables with a little board and you did a ball down the slide and both sides were equal so I must say it came up there, and it would come down there so the sides would be equal . . . the spaces in between.

From the nonverbal group came the following:

> *Henry* (7;3): I've done it before with a ball, rolling down the chute. It hits it and bounces off something. . . . The same with the flashlight. It is sort of like a ball. You aim it and it will bounce off of there. . . . If you put it slanted and the mirror was straight, it would bounce off at the same slant as the flashlight was on. If the mirror was slanted, you'd have to slant the flashlight more but it wouldn't come off in the same slant.
>
> *Rose* (8;9): It goes up even. It is balanced. It would go up like this and off like that.

In the control group the following statement was *not* typical:

> *Stephen* (7;4): You would make it so it would bounce onto the card. You look at the angle and you come down the same angle and you set the flashlight so that the angle would come down the same angle.

More characteristic were the vaguer statements:

> *Jean* (9;1): I know that if it—when the mirror is turned upward, this way, if you turn it over this way, the reflection is going to be—the more you turn this around, and more the reflection is going to be higher. . . . When the mirror is turned this way, the more you tilt the flashlight the higher it is going to be. . . . It depends on how much more tilted it is whether it is going to go higher or lower. Which way it is going to go depends on what tilt the mirror is on.
>
> *Sandra* (8;7): You must try to aim it close to the piece of paper.
>
> *Marie* (8;1): I aim it opposite the card, but not too opposite.

The motor performance of the last four children quoted was markedly different, the first two scoring near the top of the distribution, the others near the bottom.

### Verbal Performance in Relation to Motor Performance

Since verbal mediation must have occurred, motor performance was expected to be better for those showing verbal transfer. But, on the other hand, a verbal formulation may reflect, not verbal mediation, but the observations which produced better motor performance. In the control group, which had not received any prior systematic training and certainly not the verbal training, there was a relationship between verbal and motor performance (Table 6). Three of the four children who gave correct descriptions, either complete or partial, had the highest motor performances.

TABLE 6. Median Performance Scores by Verbal Score on Transfer Series B

| | Treatment | | |
|---|---|---|---|
| | Verbal | Nonverbal | Control |
| Complete rule | 8 | 10 | 13 |
| Partial correct | 9 | 11 | 13 |
| Incorrect | 8 | 8 | 7 |
| No rule | 10 | 9 | 6[*] |

[*] Significantly lower than rest of control group, at the .05 level, by Mann-Whitney test.

Furthermore, the group that offered an incorrect explanation had some slight tendency to have better performance than those who offered no explanation at all in the control group. It is not significantly different either from the group offering a correct explanation and those offering none, though these differ from each other. Under other treatment conditions this difference is attenuated, becomes nonsignificant in the nonverbal group, and disappears entirely in the verbal group. Thus, verbal training has an independent effect on verbal responses that obliterates the correlation with performance that is present without special training.

### Transfer Effects of Spontaneous Verbalization

It was proposed that the children who formulated a verbal statement themselves during training might have a higher probability of producing it again during the transfer series and that it might have a closer referential relation to the performances. In the nonverbal training group five of the 16 children did produce spontaneously a complete description corresponding to the nonverbal training. Their performance at the training session did not differ from that of those who did not formulate their observations verbally, but their verbal transfer did differ. All five gave correct descriptions of series A of the transfer test, but only two of the five gave correct descriptions in series B where the scoring was more exacting.

A somewhat comparable difference occurs in the verbally trained group between those who repeat verbatim the phrase taught and those who use their own variations during training. While there was no relation between "parroting" and performance transfer, verbalizations were somewhat related. The children who merely repeated the teacher's formulation were likely to give less satisfactory descriptions at the transfer session.

## Discussion

The experiment had been formulated in terms of a transfer and mediation design, but it was evident in the analysis that many of the children had encountered elastic collision elsewhere. One child in the control group even used the term "ricochet" in talking of reflection from a mirror.

It is clear, however, that the children who came in with less outside influence improved under training. This is the more striking because of the reduction of incidental similarities between the collision and reflection problems. The experimenter avoided similar terminology, a different experimenter was used, and fortuitous similarities in equipment were reduced as much as possible. That is, the children showed improved performance on the series B task, which is one of the most analogous to the training series, spontaneously without being reminded that the task might bear a relation to the slide and ball.

It is not clear just which aspect of the training during the first session was facilitative. Both groups given training had actual practice in aiming and adjusting the angle of approach, which as a visual-motor task would be related to the transfer series B. Both groups had been shown the path of the ball on paper and had drawn diagrams of the path to a criterion of learning. Clearly the facilitation derives from this training.

This facilitation could have arisen from arriving at a generally, but not always, correct aiming pattern. The successes with the partially usable "rules" would be the same for both sessions, and thus they would be transferable. Thus, not only the equal-angle formulation but certain others would have a fairly high proportion of success. The other rules might well receive more reinforcement because they are easier to apply, e.g., aiming midway between the projection on the board of the ball and the slide and adjusting at the second setting.

The fact that the general success of the nonverbal group is due in part to use of these alternative rules is suggested by the findings on key item B-7 on the transfer task. On this setting the children usually aimed where they would have had to aim if the mirror were more nearly horizontal. Thus, they betrayed the lack of any real systematic relation of their setting to the angle of the mirror. But in this case the children in the verbal group more frequently made a correct setting.

On the whole, the verbal group did not have as good performance scores as the nonverbal group. The difference between the total performance scores and the scores on B-7 suggests that the total scores reflect the use of any reasonably successful techniques. It is possible that the verbal training actually interfered with the discovery of alternative setting strategems. Only when the taught rule was clearly superior to the alternative rules was the verbal group superior, as on item B-7.

The verbal training used in this experiment was characterized by several special features. The terminology (N) and the information (P) were both novel, and no direct inference as to action (P → R) is immediately evident. It was difficult for the children to discern the referents of the terms; the angles of incidence and reflection are, after all, visualized and not actually visible during aiming. In addition, the implications for adjustment of aim are not directly evident. It is likely that any easier generalizations about aiming would replace the use of the more intangible equating of angles. Thus, when the verbal formulation had been taught, or even when it occurred spontaneously at the training session, it had little to do with the child's total performance in the transfer session. Even when it was available at the second session, as judged from overt verbalizations, there was no correlation with performance totals. But in the control group it can be expected that verbal formulations emerged from performance, hence the clear correlation between verbal statements and success in the control group. Verbal training had two effects: it improved performance on a key criterial task, and it altered the quality of verbal formulations. The effectiveness of the transfer on verbal behavior is the more noteworthy when it is remarked that of the 12 children in the verbal group who could, when asked, recall what they had learned about the slide, two-thirds spontaneously had transferred this description to the mirror situation.

In the nonverbal group, which had not been given names for the relationships they saw depicted, only three children could, when asked, describe what they had learned about the slide. It is small wonder, then, that their verbal performance was very little superior to that of the control group. In verbal transfer a much less complex situation obtains than in the postulated mediation of transfer of motor skills. Here there need only be a similarity of stimulus to elicit a similar verbal response. In the case of motor transfer in the nonverbal group there was also a similarity in the two overt features of the task. Only in the most sensitive task is there evidence that those with verbal mediation available to them may in fact have been using a somewhat different approach to the visual-motor task. Perhaps this verbal training oriented them differently to the slide and ball learning, and this practice with a different aiming method accounts for the differences, not second-session verbal mediation. On comparable key items in the first-session aiming the verbal group did insignificantly better. The more marked differences found at the second session cannot be due to training session practice in different visual-motor skills. We conclude, then, that the differences in key item performance at the second session are due to the occurrence of verbal mediation which increases the probability of using the solution method that is most appropriate to this item.

This experiment is relevant to curricular planning and bears certain implications for educators. It confirms the impression that pupils do spontaneously relate learning that has different concrete content. But in this experiment a deliberate effort was made to plan the training with the slide and ball so that the strategems learned would be most relevant to the reflection task. The verbal training was planned so that the terms used would apply equally in both situations. There were several routes to transfer for these children; one was verbal mediation. But the use of verbal mediation rests, like any kind of transfer, on the assumption that the situations are similar enough or, more pertinent to the teacher's power, that what the children notice about the situations has sufficient similarity, so that the children think at time II about the "generalization" that they learned at time I. In this experiment it was clear that many children did notice the similarity since they applied the terminology of the first session to the reflection problem. Yet this did not necessarily lead them to better problem solutions. Thus, the second necessary training is to link generalization to action, whatever that action may be. The great skill that children acquire in verbal learning is something of a hazard in that it is likely to reward both teacher and pupil so readily that the teacher overlooks the learning criterion she is most concerned with, which is not necessarily verbal learning alone. In this experiment we were concerned with performance as well as verbal learning. The link from speech to action was the weakest in the training of this experiment.

## Summary

Elementary scientific training might utilize the possibility that principles applicable to one concrete domain will be relevant to another and that this relevance is not merely logical and analytic but psychological as well. This psychological relationship is tested by a transfer study on third- and fourth-grade children. If the relevance is evident without explicit verbal learning of the principle which could apply to both situations, then the group without verbal training should perform as well as the group with verbal training.

When the older and more skilled children are eliminated from the comparisons, so that a larger proportion of the variance in transfer is attributable to experimental training, it appears that the children did spontaneously profit from their earlier training without the links being explicitly indicated by the teacher. In a visual-motor aiming task using a flashlight and mirror to reflect on a target, children performed better who had practiced earlier with a ball rebounding from a board.

The group with additional verbal training did not show as much facilitation in the total motor performance as the group with only motor training. But on a key question which could not be solved as easily by aiming strategems other than the one taught, the verbal group was superior. The verbal group also was superior in the descriptions they offered of the principle involved in reflection.

Statements about the reflection principle could either influence performance or be derived from it. In the control group there was a correlation between performance and verbal statements, but not in the other groups. One reason for the lack of relationship in the other groups may be the lack of direct implication as to performance strategem in the verbal statement that was taught. Thus, a person might recall the statement, but his performance might not thereby be improved. On the whole, the study confirms the theory that verbal training facilitates transfer, if several necessary conditions are present.

## REFERENCES

1. Arnoult, M. D. Stimulus predifferentiation: Some generalizations and hypotheses. *Psychol. Bull.*, 1957, **54**, 339–350.
2. Bastian, J. R. Response chaining in verbal transfer. Studies on the role of language in behavior. Univer. of Minn., under contract number N8 onr-66216 with the Office of Naval Research, 1957.
3. Battig, W. F. The effect of kinesthetic, verbal, and visual cues on the acquisition of a lever-positioning skill. *J. exp. Psychol.*, 1945, **47**, 371–380.
4. Dietze, Doris. The facilitating effect of words on discrimination and generalization. *J. exp. Psychol.*, 1955, **50**, 255–260.
5. Gagné, R. M., & Baker, K. E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
6. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
7. Hake, H. H., & Eriksen, C. W. Effect of number of permissible response categories on learning of a constant number of visual stimuli. *J. exp. Psychol.*, 1955, **50**, 161–167.
8. Howes, D. H., & Solomon, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.

9. Jeffrey, W. E. The role of verbal and nonverbal responses in mediating an instrumental act. *J. exp. Psychol.*, 1953, **45**, 327–333.
10. McGinnies, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 224–251.
11. Murdock, B. B. Transfer designs and formulas. *Psychol. Bull.*, 1957, **54**, 313–326.
12. Norcross, Kathryn J., & Spiker, C. C. Effects of mediated associations on transfer in paired-associate learning. *J. exp. Psychol.*, 1958, **55**, 129–134.
13. Porter, L. W., & Duncan, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.
14. Robinson, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, 1955, **49**, 112–114.
15. Russell, W. A., & Storms, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.
16. Underwood, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first-list learning. *J. exp. Psychol.*, 1951, **42**, 44–53.
17. Vanderplas, J. M. Transfer of training and its relation to perceptual learning and recognition. *Psychol. Rev.*, 1958, **65**, 375–385.
18. Young, R. K., & Underwood, B. J. Transfer in verbal materials with dissimilar stimuli and response similarity varied. *J. exp. Psychol.*, 1954, **47**, 153–159.

# Part VII

## *Recent Innovations in Teaching-Learning*

ONE OF THE teacher's major responsibilities is to arrange conditions in the classroom setting so that they will be likely to promote efficient student learning. Several recent developments, suggesting new ways both of viewing and establishing a setting and context for learning, have aroused considerable interest among educators and psychologists. Some of these developments have taken the form of new kinds of administrative organization for instruction. Others are the result of technological advances. Part VII: Recent Innovations in Teaching-Learning contains six selections intended to familiarize the reader with teaching machines and programed instruction—with educational television, language laboratories, team teaching, nongraded programs, and core programs.

Charles Morrill reviews, in the first selection, progress in the area of teaching machines and programed instruction. At the time this article was published Morrill was a staff psychologist with the MITRE Corporation. His review highlights the flurry of activity currently surrounding self-instructional materials and devices.

Discussion of another technological innovation, educational television, is the substance of the following selection, by Walter Wittich. Wittich, professor of education at the University of Hawaii, predicts continuous development of effective educational television.

Milton and Helen Shane describe, in reading number 40, the advantages afforded by language laboratories in foreign language teaching. Their article is based on their own experiences with language laboratories, at both the high-school and college levels, at the George Peabody College for Teachers, in Nashville, Tennessee. The language-laboratory concept represents another technological development with important instructional implications.

An assessment of "team teaching," one of the new forms of organization for instruction, is provided by Harold Drummond in reading number 41. Drummond, Chairman of the Department of Elementary Education, the University of New Mexico, calls for more sophisticated research designs in team-teaching experimentation to enable us to control more carefully the many variables involved.

Hugh Perkins, professor of education at the Universsity of Maryland, assesses the progress of nongraded programs, in reading number 42. According to Perkins, such programs provide "a flexible grouping and organization of curriculum content that enable children of varying abilities and rates of maturing to experience continuous progress in learning."

*In the final selection Gordon Vars discusses block time, unified studies, and core programs. Vars, professor of education at Cornell University, holds that experience has demonstrated the potential advantages of these three kinds of programs, and emphasizes the key role of administrative leadership in their development.*

# 38

# Teaching Machines: A Review

## — *Charles S. Morrill*

*The use of programed instructional materials and of other auto-instructional devices has aroused considerable attention among educators, who are faced with the need of administering education rapidly to large numbers of students. In this article the author presents background material to the development of teaching machines. Both the history and theory of the teaching machine movement are discussed, together with current research in the field. Specific questions for further investigation are raised; and, finally, problems of application are discussed. The present reading should provide a good background to the understanding of self-instructional materials and devices.*

THE SAME forces which have characterized the evolution of general educational practices are inherent to the history of the new science of automated teaching. As a result of the expansion and multiplying complexities of political, economic, and social interests, there developed an ever increasing need for the rapid education of large numbers of people. New educational objectives demanded new methods of instruction, and the history of education is marked by many diverse attempts at establishing more efficient teaching procedures. Once again teaching methods must be re-evaluated. Rigid adherence to the principle of personal teacher-student relationships no longer seems feasible—an instructional system more appropriate for present-day needs must be established. It is probable that the use of auto-

Reprinted from *Psychological Bulletin*, 1961, **58**, 363–375, by permission of the author and the American Psychological Association.

mated teaching devices can fill this need in the method of education. As Corrigan (1959) has suggested:

> the automated teaching method has grown out of a pressing need. This need has been created by a twofold technical training problem. As advances in science and technology have been made, there has been an ever increasing demand for well-trained instructors; at the same time the availability of these trained persons has been diminishing. This situation is aggravated further by the increased scope and complexity of subjects, and the ever increasing ratio between number of instructors and students (p. 24).

## Current Trends in Automated Teaching Machines

Current interest in the area of automated teaching machines is well illustrated by the simple index of frequency-per-year of published teaching machine articles. Fry, Bryan, and Rigney (1960) report that for the years prior to 1948 there are only 6 references, whereas through 1959 there were more than 50 reports published.

The grandfather of automated teaching machines is Sydney L. Pressey (1926, 1927), who designed machines for automated teaching during the mid-1920s. His first device was exhibited and described at the American Psychological Association (APA) meetings in 1924; an improved device was exhibited in 1925 at the APA meetings. Both forms of the apparatus automatically performed simultaneous administration and scoring of a test and taught informational and drill material. Pressey's device, about the size of a portable typewriter, presented material to the subject via a small window. Four keys were located alongside the apparatus. If the student activated a key corresponding to the correct answer, the machine advanced to the next item. If his response was incorrect, the machine scored an error and did not advance to the next item until the correct answer was chosen. The capacity of the drum was 30 two-line typewritten items; the paper on which the questions appeared was carried as in a typewriter.

In 1927, Pressey summarized his efforts as follows:

> The paper reports an effort to develop an apparatus for teaching drill material which (a) should keep each question or problem before the learner until he finds the correct answer, (b) should inform him at once regarding the correctness of each response he makes, (c) should continue to put the subject through the series of questions until the entire lesson has been learned, but (d) should eliminate each question from consideration as the correct answer for it has been mastered (p. 552).

In 1930, Peterson devised a self-scoring, immediate feedback device. The Chemo Card, as this device was later called, utilized the technique of multiple choice. A special ink was used by the student in marking his answer. The mark appeared red if the answer was incorrect; a dark color resulted if the answer was correct. Although Pressey's notions and the Chemo Card might have stimulated an interest in automated teaching techniques in the twenties, educators and researchers obviously were not at that time ready for this advanced concept of teaching. Automated teaching did not take hold.

In 1932, Pressey published an article describing a kind of answer sheet which could be scored by an automatic scoring device. This apparatus recorded errors by item, and thus provided the instructor with clues as to what questions needed further instruction. In 1934, Little experimented with this device as well as with the device originated by Pressey in 1926. His results favored the use of automated devices in contrast to regular classroom techniques.

The next appearance of automated teaching literature came a considerable number of years later. During World War II, the Automatic Rater was used by the Navy for training. This device projected a question on a small screen; the subject's response consisted of pushing one of five buttons.

In 1950, Pressey described a new automated device called the Punchboard. Multiple-choice questions were presented to the student. The key answer sheet inside the Punchboard contained holes opposite the correct answers only. If the answer was correct, the student's pencil penetrated deeply; if incorrect, the pencil did not penetrate the paper significantly. Angell and Troyer in 1948 and Angell in 1949 reported the results of using the Punchboard. Both studies suggested the superiority of this method over traditional classroom procedures.

In 1954, Skinner published "The Science of Learning and the Art of Teaching," which provided the basis for the development of his teaching machines. In this article, he stressed the importance of reinforcement in teaching and suggested teaching machines as a method of providing this needed reinforcement for the learner.

Reports concerning the Subject-Matter Trainer began to appear in 1955 (Besnard, Briggs, Mursch, & Walker, 1955; Besnard, Briggs, & Walker, 1955). This electromechanical device is a large multiple-choice machine used essentially for training and testing in the identification of components and in general verbal subject matter. Extensive research has been done with this device because of its considerable flexibility, i.e., it allows several modes of operation for self-instruction: variety of programed subject matter, drop-out feature after items have been mastered, etc.

The Pull-Tab, used experimentally by Bryan and Rigney in 1956, was a device in which the subject received not only a "right" or "wrong" indication after his choice but also a somewhat detailed explanation of "why" a response was incorrect. In 1949, Briggs had found in experimenting with the Punchboard that learning is significantly enhanced by immediate knowledge of results. Bryan and Rigney's data illustrated that the combination of immediate knowledge of results plus explanation, if the student is in error, produced significantly higher scores on a criterion test than if no explanation had been given. The importance of this research from a historical point of view is that it investigated immediate knowledge of results as a factor existing on a continuum with varying degrees of effect. Up to this point any comparison involving the effectiveness of teaching machines had been one between classroom instruction and the "new" machine under consideration. In Briggs' and in Bryan and Rigney's research, however, we see the beginning of a concern, to become greater in the next few years, with the possible effects of specific variables and their interactions on learning.

The years 1957–58 mark the beginning of the period in which resurgent interest in teaching machines was initiated. Ramo's arguments (1957) reopened the consideration of automated techniques for classroom use. His article served as one of the more forceful attempts to alert educators to the needs and requirements for automated techniques in education. Skinner's continued interest (1958) served as the major catalyst in this area. In his article, he reviewed earlier attempts to stimulate interest in teaching machines and further explained that the learning process was now better understood and that this increased sophistication would be reflected in teaching machine technology. Skinner suggested that the most appropriate teaching machine would be that which permits the student to *compose* his response rather than to select it from a set of alternatives. On the basis of this philosophy and in conjunction with other principles of learning theory to which Skinner adheres, he designed a teaching machine with the following characteristics. The questions, printed on a disk, are presented to the student through a window. The student's response is written on a paper tape, which is advanced under a transparent cover when the student lifts a lever. At this point the correct answer appears in the window. If the student is correct, he activates the lever in one manner, which eliminates the item from the next sequence. If he is incorrect, the lever is activated in a different manner, thus retaining the item in the next sequence.

Holland (1960), a co-worker of Skinner's, has suggested several well-known learning principles that should be applied to teaching machine technology: immediate reinforcement for correct answers is a must, learned behavior is possible only when it is *emitted* and reinforced, gradual progression (i.e., small steps in learning sequences and reducing wrong answers) is necessary to establish complex repertoires, gradual withdrawal (fading or vanishing) of stimulus support is effective, it is necessary to control the student's observing and echoic behavior and to train for discrimination, the student should write his response. The Skinner machine does in fact employ these principles.

Ferster and Sapon (1958) described the Cardboard Mask, a most simple teaching machine which employs the principles which Skinner and Holland outline so clearly. This device is a cardboard folder containing mimeographed material which is presented one line at a time. The student, after writing his response on a separate sheet of paper, advances the paper in the mask, thereby exposing the correct response.

In 1958, a number of investigators interested in teaching machines recommended that the programed material be a function of the student's response. This idea suggests that a "wrong" response may not necessarily be negative reinforcement and that both the "right" and "wrong" responses should modify the program. Rath and Anderson (1958) and Rath, Anderson, and Brainerd (1959) have suggested the use of a digital computer which automatically adjusts problem difficulty as a function of the response. Crowder's (1955, 1958, 1959a) concept of "intrinsic programming" permits the response to alter the programing sequence.

During the last few years, researchers have been focusing their attention on investigating many of the variables which are pertinent to the design and use of teaching machines. The seemingly simple task of defining a teaching machine has been a serious problem to many authors (Day, 1959; Silberman, 1959; Weimer, 1958). Some definitions have made more extensive demands on teaching devices than others. Learning theorists (Kendler, 1959; Porter, 1958; Skinner, 1957; Spence, 1959; Zeaman, 1959) are now most outspoken concerning the application of theoretical concepts to teaching machine technology. Transfer of training, mediational processes, reinforcement, motivation, conditioning, symbolic processes, and language structure are but a few of these areas of interest.

There are indeed many other variables about which there is a divergence of opinion and about which experimental evidence is completely lacking or controversial. The reports of Skinner (1958), Israel (1958), Coulson and Silberman (1960), Fry (1959), and Stephens (1953) are all focused, at least in part, on questions related to response modes, e.g., multiple choice, construction of the response, responses with reinforcement, etc. Briggs, Plashinski, and Jones (1955) investigated self-paced vs. automatically paced machines. The importance of motivation in connection with teaching machines has been explored by Holland (unpublished), Mayer and Westfield (1958), and Mager (1959).

Essentially, the history of automated teaching is short—it started in the mid-twenties and was strenuously reactived by the appearance of Skinner's 1958 article. Empirical investigations of many important issues in this field are just now beginning to appear. However, the necessity of developing automated teaching methods has been evident for many years.

## General Problem Areas

### Definition

As in any new field, the first problem is one of definition. What is a teaching machine? Silberman (1959) says that a teaching device consists of four units: an input unit, an output unit, a storage unit, and a control unit. As such, this definition includes a broad category of devices, from the most simple to the most complex. Weimer (1958) goes beyond the device itself, stating that a teaching machine must present information to the student as well as test the student by means of a controlled feedback loop. Crowder (1960) insists that a teaching machine

> must in some way incorporate two-way communication. That is, the student must respond to the information presented by the machine, and the machine must in turn recognize the nature of the student's response and behave appropriately (p. 12).

Perhaps the most inclusive definition is one given by Day (1959):

> A teaching machine is a mechanical device designed to present a particular body of information to the student. . . . Teaching machines differ from all other teaching devices and aids in that they require the active participation of the learner at every step (p. 591).

Although the emphasis in some of the above concepts is different, together they give a rather complete description and, if you will, definition.

### Programing

The programing of subject matter for teaching machines is the most extensive and difficult problem in this new technology. Beck (1959) describes specific concepts which he thinks appropriate for programing a Skinner-type machine:

> A student's responses may be restricted and guided in a great number of ways. These range from all types of hints . . . to simply presenting the response which it is desired a student acquire (p. 55).

Carr (1959) discusses in some detail the importance of programing in terms of learning efficiency and retention. Much of what he says remains open for empirical verification. Rothkopf (1960) has suggested that the development of programed instruction suffers from two difficulties: a weak rational basis for program writing and inadequate subject-matter knowledge among program writers.

The extent to which any initial program needs revision is perhaps exemplified by the program in Harvard's course Natural Sciences 114. Holland points out that the first program of materials included 48 disks, each containing 29 frames, whereas a revision and extension of the program the following year included 60 disks of 29 frames each. Holland's objective was to extend the program and decrease the number of student errors. Crowder's (1960) programing objectives are different from Holland's. He states:

> By means of "intrinsic programming" it [the program] recognizes student errors as they occur and corrects them before they can impede understanding of subsequent material or adversely affect motivation (p. 12).

Crowder considers it almost impossible to write a program which completely avoids error, and therefore he structures the program requirements on the probability of error. When an error is made, the next presentation explains the subject's mistake. Depending on the nature of the error and when it occurs, the subject may either return to the original question or enter a program of correctional material.

Another concept for programing is known as *branching* (Bryan & Rigney, 1959). Through branching, many possible routes are provided through which the subject can proceed, depending on the response. The subjects are allowed to skip certain material if they have demonstrated a knowledge of it. One study (Coulson & Silberman, 1960) suggests that under branching conditions subjects require less training time than under non-branching conditions; however, results on the criterion test were not significantly different.

For certain kinds of subject matter, *vanishing* is still another concept for programing (Skinner, 1958). A complete or nearly complete stimulus is presented to the subject. Subsequent frames gradually omit part of the stimulus until all of it is removed. The subject is then required to reconstruct the stimulus.

To program verbal learning sequences, Homme and Glaser (1959) suggest the Ruleg. With this method, the written program states a rule and provides examples for this rule. In each case, either the rule or the example is incomplete, requiring the subject to complete it.

In a recent study Silverman (1960b) investigated methods of presenting verbal material for use in teaching machines. He recommended that further research involving the design and use of teaching machines should take into consideration the possible use of context cues as a means of facilitating serial rote learning. At the same time, however, he stated that continuous use of context cues as ancillary prompts should be avoided, since such prompts can interfere with learning.

The optimum size of steps and the organization of the programed material are two formidable problems. Skinner (1958) states:

> Each step must be so small that it can always be taken, yet in taking it the student moves somewhat closer to fully competent behavior (p. 2).

In order to determine the value of steps in a program, Gavurin and Donahue (1961) investigated the effects of the organization of the programed material on retention and rate of learning. They state that the assumption that optimum teaching machine programs are those in which items are presented in a logical sequence has been validated for acquisition but not retention. The results of the study carried out by Coulson and Silberman (1959) indicated that small steps were more time consuming but resulted in statistically significant higher test scores on one of the criterion tests. Pressey (1959) in principle disagrees with Skinner's notions of short and easy steps, and he strongly suggests an experimental investigation of this question. Both rate of learning and retention (recall or recognition) are of critical concern.

The above discussion suggests several areas which are directly applicable to programing and which are under investigation and/or need further experimentation. Indeed, there are a number of unanswered questions in the programing complex, some of which have been suggested by Galanter (1959):

1. What is the correct order of presentation of material?
2. Is there an optimum number of errors that should be made?
3. How far apart (in some sense) should adjacent items be spaced?
4. Is experimentally controlled pacing more effective (in some sense) than self-pacing?
5. Is one program equally effective for all students?
6. What are the effects of using different programing techniques (branching, intrinsic programing, vanishing) in various subject-matter areas?
7. What criteria are most appropriate in the evaluation of student learning?

These questions are but a few of the intriguing and complex problems facing investigators in the new field of programing material for teaching machines. Answers to these questions will help not only the educator but also the engineer who is concerned with writing adequate specifications for the construction of teaching machines.

## Response Mode

The kind of response that should be given by a subject has been a controversial question in the teaching machine field. Pressey's original machine (1926) required the subject to press a lever corresponding to his choice of answer. The format of the answers was multiple-choice. Skinner

(1958) emphasized the necessity of having the subject *compose* (construct) the response. Skinner states:

> One reason for this is that we want him to recall rather than recognize—to make a response as well as see that it is right. Another reason is that effective multiple-choice material must contain plausible wrong responses, which are out of place in the delicate process of "shaping" behavior because they strengthen unwanted forms (p. 2).

Coulson and Silberman (1960) investigated this question of multiple-choice vs. constructed response by using *simulated* teaching machines—human beings were used instead of automatic control mechanisms. Their results indicated that the multiple-choice response mode required significantly less time than the constructed response mode and that no significant difference was obtained between response modes on the criterion test. Further, they reported that no significant differences were obtained among the experimental groups on the multiple-choice criterion subtest or on the total (multiple-choice plus constructed response) criterion test. Fry (1959) has discussed this response-mode question along with other variables, and he has carried out extensive research concerning constructed vs. multiple-choice response modes. The results of his study favor the use of constructed response when recall is the objective of the learning.

In addition to the basic controversy (which needs much more investigation) between multiple-choice and constructed responses, there are several "variations on the theme" which are evident. Stephens (1953) has recommended that every wrong answer in a multiple-choice question appear as a correct choice for another item. He calls this program "inside alternatives." His data indicate that there was no difference between control and experimental groups on a criterion test using either nonsense syllables or Russian unless each right choice appeared as a wrong alternative for the three subsequent items. The use of prompts in general has been shown to be an effective technique in automated teaching (Cook, 1958; Cook & Kendler, 1956; Cook & Spitzer, 1960).

Using learning booklets, Goldbeck (1960) investigated the effect of response mode and learning material difficulty on automated instruction. The three response modes used were: overt response (the subject was required to construct a written response), covert response (the subject was permitted to think of a response), and implicit response (the subject read the response which was underlined). Goldbeck states:

> Learning efficiency scores, obtained by dividing quiz scores by learning time, showed that the implicit (reading) response condition produced significantly more efficient learning than the overt response condition. The covert response condition fell between the other conditions in learning efficiency (p. 25).

Concerning quiz-score results, the overt response group

> performed significantly poorer than the other response mode groups at the easy level of difficulty. Performance of the overt response group improved significantly at the intermediate difficulty level to the extent that it exceeded the performance of all other groups (pp. 25–26).

Goldbeck concludes that

> doubt is cast upon the assumption that the best learning is achieved by use of easy items and requiring written constructed responses (p. 26).

To the author's knowledge, the use of an oral response in conjunction with the Skinner teaching machine and its effect on learning rate and retention have not been reported in the literature. Furthermore, the importance of response mode as a function of reinforcement must be specified. Israel (1958) has suggested that natural and artificial reinforcement may affect the subjects' learning. A most comprehensive analysis of response-mode and feedback factors has been reported by Goldbeck and Briggs (1960).

The general area of reinforcement suggests problems related to the drop-out feature of teaching machines. Pressey's (1927) original machine dropped items after the correct answer had been given twice. Skinner's machines at the Harvard Psychological Laboratory also have the drop-out feature, although the commercially available machines based on Skinner's design do not incorporate this feature. With reference to a study carried out at Harvard, Holland (unpublished) reported significantly superior performance when the drop-out feature was used.

If items are dropped, the sequence of items is of course changed. How important is the sequence? If items should be dropped, by what criterion of learning can one justify omitting an item from the sequence? If items are not dropped and the criterion for the learning procedure is a complete run (i.e., once through the sequence without error), what is the effect upon retention? Being correct is positive reinforcement; thus, some items under these circumstances will receive a greater amount of positive reinforcement than others. What would be the effect of additional reinforcements with or without drop-out? Again, a plethora of problems and a paucity of answers!

Response time, another important variable, has been investigated by Briggs, Plashinski, and Jones (1955). Their study suggests that there is no difference between self-paced and automatically paced programs as determiners of response time. However, the problem of pacing for individual items is still a recent one and needs further research. Another aspect of response time—the distribution of practice—has been studied extensively since Ebbinghaus' investigation in 1885. For example, Holland (unpublished) states that in an experiment at Harvard "a few students completed all the disks in a small number of long sessions while others worked in many short sessions. . . . Apparently the way practice was distributed made little difference" (p. 4). Nevertheless, the distribution of practice, like the problem of pacing, is yet a subject of controversy, with most investigations favoring some form of distributed practice (Hovland, 1951).

The above section outlines briefly some of the major problems associated with the variables affecting response mode. Although some of the variables have already been investigated, these and others, together with their interactions, need further research.

### Knowledge of Results

There are many peripheral problems related to teaching machines, one of which is the effect of immediate knowledge of results on learning. Angell (1949), using a multiple-choice punchboard technique, found that "learning is significantly enhanced by immediate knowledge of results." Briggs (1949), also using the Punchboard, confirmed these results. Bryan and Rigney (1956) noted superior performance when subjects were given knowledge of results, specifically, an explanation if the answer was incorrect. This last study was later expanded by Bryan, Rigney, and Van Horn (1957), who investigated differences between three kinds of explanation given for incorrect response. None of the three types of explanation proved to be superior in teaching the subjects. Because of their controvertible results, the above studies demonstrate that, although immediate knowledge of results appears to be effective in the learning process, this problem contains many facets which need more empirical data.

### Motivation

One of the many reasons given for the effectiveness of teaching machines is that the student's motivation is increased. Psychologists and educators have realized for some time that the motivation variable ranks very high among those variables pertinent to learning. In 1958 and 1959, Holland surveyed the use of the teaching machine in classes at Harvard. He found that most students felt that they would have gotten less out of the course if the machines had not been used, that most students preferred to have machines used for part of the course, and finally that most students felt that the teaching machine was used by the instructor "to teach me as much as possible with a given expenditure of my time and effort." During a field tryout of the Subject-Matter Trainer in the Semiautomatic Ground Environment System, Mayer and Westfield (1958) observed that "motivation to work with the trainer is high." The supervisory as well as the operational personnel encouraged the use of this training technique.

Mager (1959) suggests that motivation and interest are a function of the percentage of correct responses. He observed that in two young subjects negative feelings for learning mathematics in the usual classroom situation did not transfer to learning mathematics by means of a teaching machine. The cause of this phenomenon is perhaps best explained by the subjects' statement that, because they were able to understand the programmed material, it did not seem to be mathematics at all. This interesting relationship between comprehension and motivation needs further investigation.

## Equipment

There are many inexpensive models of teaching machines which will soon hit the consumer market. For much of this equipment, there is very little experimental evidence which supports the various designs. As previously pointed out, Holland has collected data which support the efficiency of the drop-out feature in a teaching machine; yet commercial models presently available do not incorporate this feature, presumably because of its high cost. Generally, it seems that production is now and will continue to be out of phase with much of the research which has provided necessary teaching machine specifications. Moreover, because of their expense, it is likely that some very important features will be omitted in manufacture.

The methods of displaying programed material, another unexplored problem area, must be investigated so as to provide the design engineer with requirements based on empirical findings. The display problem is less acute, perhaps, with material for the elementary school than it is with programs designed to teach maintenance procedures and aspects of the biological sciences.

The use of computer controlled teaching machines has been recommended by many authors (Coulson & Silberman, 1959; Skinner, 1958). Utilizing a central computer, with many programs capable of adapting to individual needs and of providing stimulus materials to 50 or more students simultaneously, is a feasible notion for large-scale training programs. With a computer, the display problem again becomes a major issue. Training in pattern recognition, information handling, and display interpretation are but a few appropriate areas which should be studied. The alternate modes of presentation become more extensive as computer capacity increases. In the case of certain kinds of subject matter, a computer generated, pictorial display of information may be a more effective presentation than other display techniques. Future research must solve these problems in equipment design.

## Teaching Machines and Other Techniques

The use of automated teaching devices may be optimized, perhaps, if there is a proper balance between this technique and other compatible teaching methods. What percentage of a course should be machine taught? What subject matter is best suited to automated devices? If classroom courses were as carefully and thoughtfully programed as some of the programs currently being prepared for teaching machines, might some of the advantages of machines diminish? Perhaps some of the apparent advantages of teaching machines are no more than methods of illustrating correctable classroom techniques! It might well be that the instructor's enthusiasm and inspiration, a factor supposedly dominant in higher education, is vital in mastering a particular subject-matter area. Will creativity in certain students be harmed by extensive education via the machine? Again, consideration of the use of a teaching machine, the subject matter, the program, the level of education, and the techniques used in combination with the teaching machine provide a fertile field for experimentation. As of now, questions in this area remain unanswered. Silverman (1960a) has presented an excellent, detailed discussion of problems inherent in this new technology of automated teaching and the current trends in the field.

### Problems of Application

The most obvious problems in the attempt to use automated teaching techniques have been outlined in the previous section. There is still much of the unknown associated with techniques, machines, programing, etc., to be eliminated before a direct solution to a particular training problem can be specified. Many alternatives exist, the best of which has not yet been determined. In addition to these voids, there is a serious lack of definition in the objectives of many training programs.

What is the objective of a particular automated course or program? From a pragmatic point of view, what are the criteria by which a specific educational program can be evaluated? For example, the objectives might range from the teaching of rote tasks to the presentation of more abstract material. Needless to say, the techniques for both teaching and evaluating learning could be substantially different in each case. The purpose of teaching, the objective of an educational program, must be initially defined. Only then will the concepts *learning* and *teaching* be meaningful in a particular context.

After definition, the next step is to determine what subject matter will provide the student with the necessary information. It is at this point that the major pitfall in education is likely to appear. Even though many training programs do not have a defined objective, their course content is nonetheless prescribed, and the text and/or materials used in previous, nonautomated courses become the prime source of material for an automated teaching program. To program an automated teaching machine with presently available materials might well result only in a more efficient method of teaching the wrong material!

The third step requires decisions in the selection of appropriate teaching techniques. Answers to questions involving programing, choice of teaching machine, learning procedures, pacing, and response modes are still not known.

The fourth and last step requires an evaluation of the selected automated teaching method in terms of the originally established objectives. Conventional methods of instruction should be compared with the innovative methods by means of a specific set of criteria, e.g., in terms of training time, job performance, retention of learned information, etc.

The questions confronting the researcher in teaching machine technology are one example of the broader questions of man-machine interrelation. Data pertinent to the principles of human engineering, the optimum man-machine interaction, the degree to which the machine can perform functions formerly allocated to man, and the appropriate allocation of functions between man and machine will be provided by a research program investigating teaching machines. Inadequate attention to any of the above-mentioned steps will result in failure to provide the needed answers in a field which may increase training effectiveness and reduce training costs.

## REFERENCES

Angell, G. W. The effect of immediate knowledge of quiz results on final examination scores in freshman chemistry. *J. educ. Res.*, 1949, 42, 391–394.

Angell, G. W., & Troyer, M. E. A new self-scoring test device for improving instruction. *Sch. Soc.*, 1948, 67, 84–85.

Beck, J. On some methods of programming. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art.* New York: John Wiley, 1959. pp. 55–62.

Besnard, G. G., Briggs, L. J., Mursch, G. A., & Walker, E. S. Development of the subject-matter trainer. *USAF Personnel Train. Res. Cent. tech. Memo.*, 1955, No. ASPRL-TM-55-7.

Besnard, G. G., Briggs, L. J., & Walker, E. S. The improved subject-matter trainer. *USAF Personnel Train. Res. Cent. tech. Memo.*, 1955, No. ASPRL-TM-55-11.

Briggs, L J. The development and appraisal of special procedures for superior students and an analysis of the effects of knowledge of results. *Abstr. Doctoral Dissertations,* Ohio State Univer., 1949, No. 58.

Briggs, L. J., Plashinski, D., & Jones, D. L. Self-pacing versus automatic pacing of practice on the subject-matter trainer. *USAF Personnel Train. Res. Cent. lab. Note,* 1955, No. ASPRL-LN-55-8.

Bryan, G. L., & Rigney, J. W. An evaluation of a method for ship-board training in operations knowledge. *Univer. Sth. Calif. Electronics Personnel Res. Group tech. Rep.*, 1956, No. 18.

Bryan, G. L., & Rigney, J. W. Current trends in automated tutoring and their implications for naval technical training. *Univer. Sth. Calif. Dept. Psychol. tech. Rep.*, 1959, No. 29.

Bryan, G. L., Rigney, J. W., & Van Horn, C. An evaluation of three types of information for supplementing knowledge of results in a training technique. *Univer. Sth. Calif., Electronics Personnel Res. Group tech. Rep.*, 1957, No. 19.

Carr, W. J. Self-instructional devices: A review of current concepts. *USAF WADC tech. Rep.*, 1959, No. 59-503.

Cook, J. O. Supplementary report: Processes underlying learning a single paired-associate item. *J. exp. Psychol.*, 1958, **56**, 455.

Cook, J. O., & Kendler, T. S. A theoretical model to explain some paired-associate learning data. In G. Finch & F. Cameron (Eds.), *Symposium on Air Force human engineering, personnel, and training research*. Washington, D.C.: National Academy of Sciences–National Research Council, 1956. Pp. 90–98.

Cook, J. O., & Spitzer, M. E. Supplementary report: Prompting versus confirmation in paired-associate learning. *J. exp. Psychol.*, 1960, **59**, 275–276.

Corrigan, R. E. Automated teaching methods. *Automated teach. Bull.*, 1959, **1**(2), 23–30.

Coulson, J. E., & Silberman, H. F. Proposal for extension of automated teaching projects. *Sys. Develpm. Corp. field Note*, 1959.

Coulson, J. E., & Silberman, H. F. Effects of three variables in a teaching machine. *J. educ. Psychol.*, 1960, **51**, 135–143.

Crowder, N. A. The concept of automatic tutoring. *USAF Personnel Train. Res. Cent. organizational Pap.*, 1955.

Crowder, N. A. *An automatic tutoring book on number systems*. Vol. 1. Timonium, Md.: Hover Electronics Co., 1958.

Crowder, N. A. Automatic tutoring by means of intrinsic programming. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art*. New York: John Wiley, 1959. Pp. 109–116.

Crowder, N. A. The "tutor." *J. Amer. Soc. Train. Dir.*, 1960, **14**(5), 12–17.

Day, J. H. Teaching machines. *J. chem. Educ.*, 1959, **36**, 591–595.

Ferster, C. B., & Sapon, S. M. An application of recent developments in psychology to the teaching of German. *Harvard educ. Rev.*, 1958, **28**, 58–69.

Fry, E. B. Teaching machine dichotomy: Skinner versus Pressey. Paper presented at American Psychological Association, Cincinnati, September 1959.

Fry, E. B., Bryan, G. L., & Rigney, J. W. Teaching machines: An annotated bibliography. *Audio-Visual commun. Rev.*, 1960, **8**, Suppl. 1, 1–80.

Galanter, E. H. The ideal teacher. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art*. New York: John Wiley, 1959. Pp. 1–11.

Gavurin, E. I., & Donahue, V. M. Logical sequence and random sequence. *Automated teach. Bull.*, 1961, **1**(4), 3–9.

Goldbeck, R. A. The effect of response mode and learning material difficulty on automated instruction. *Amer. Inst. Res. tech. Rep.*, 1960, No. AIR-328-60-IR-124.

Goldbeck, R. A., & Briggs, L. J. An analysis of response mode and feedback factors in automated instruction. *Amer. Inst. Res. tech. Rep.*, 1960, No. AIR-328-60-IR-133.

Holland, J. G. Teaching machines: An application of principles from the laboratory. In, Proceedings of the Educational Testing Service Invitational Conference, October 1959, *The impact of testing on the educational process.* Princeton, N. J.: Educational Testing Service, 1960.

Homme, L. E., & Glaser, R. Problems in programing verbal learning sequences. Paper presented in symposium on research issues in study of human learning raised by developments in automated teaching methods, American Psychological Association, Cincinnati, September 1959.

Hovland, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: John Wiley, 1951. Pp. 613–689.

Israel, M. L. Skinnerian psychology and educational redesign. Paper read in symposium, American Psychological Association, Washington, D.C., September 1958.

Kendler, H. H. Teaching machines and psychological theory. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art.* New York: John Wiley, 1959. Pp. 177–185.

Little, J. K. Results of use of machines for testing and for drill upon learning in educational psychology. *J. exp. Educ.*, 1934, 3, 45–49.

Mager, R. F. Preliminary studies in automated teaching. *IRE Trans. Educ.*, 1959, **E-2**, 104–107.

Mayer, S. R., & Westfield, R. L. A field tryout of a teaching machine for training in SAGE operations. *USAF Cambridge Res. Cent. tech. Memo.*, 1958, No. OAL-TM-58-16.

Peterson, J. C. A new device for teaching, testing, and research in learning. *Trans. Kans. Acad. Sci.*, 1930, 33, 41–47.

Porter, D. Teaching machines. *Harvard Grad. Sch. Educ. Ass. Bull.*, 1958, 3(1), 1–5.

Pressey, S. L. A simple apparatus which gives tests and scores–and teaches. *Sch. Soc.*, 1926, 23, 373–376.

Pressey, S. L. A machine for automatic teaching of drill material. *Sch. Soc.*, 1927, 25, 549–552.

Pressey, S. L. A third and fourth contribution toward the coming "industrial revolution" in education. *Sch. Soc.*, 1932, 36, 668–672.

Pressey, S. L. Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *J. Psychol.*, 1950, 29, 417–447.

Pressey, S. L. Certain major psycho-educational issues appearing in the conference on teaching machines. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art.* New York: John Wiley, 1959. Pp. 187–198.

Ramo, S. A new technique of education. *Engng. sci. Mon.*, 1957, 21 (October), 17–22.

Rath, G. J., & Anderson, Nancy S. The IBM research center teaching machine project: I. The teaching of binary arithmetic. II. The simulation of a binary arithmetic teaching machine on the IBM 650. Paper presented at USAF Office of Scientific Research symposium on teaching machines, Univer. of Pennsylvania, December 8–9, 1958.

Rath, G. J., Anderson, Nancy S., & Brainerd, R. C. The IBM research center teaching machine project. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art.* New York: John Wiley, 1959. Pp. 117–130.

Rothkopf, E. Z. A do-it-yourself kit for programmed instruction. *Teach. Coll. Rec.*, 1960, **62**, 195–201.

Silberman, H. F. Introductory description of teaching machines (physical characteristics). Paper read in symposium on automated teaching, Western Psychol. Ass., San Diego, Calif., April 1959. (Abstract)

Silverman, R. E. Automated teaching: A review of theory and research. USN Training Device Center, 1960. (ASTIA AD-241 283) (a)

Silverman, R. E. The use of context cues in teaching machines. USN Training Device Center, 1960. (ASTIA AD-238 777) (b)

Skinner, B. F. The science of learning and the art of teaching. *Harvard educ. Rev.*, 1954, **24**, 86–97.

Skinner, B. F. *Verbal behavior.* New York: Appleton-Century-Crofts, 1957.

Skinner, B. F. Teaching machines. *Science*, 1958, **128**, 969–977.

Spence, K. W. The relation of learning theory to the technology of education. *Harvard educ. Rev.*, 1959, **29**, 84–95.

Stephens, A. L. Certain special factors involved in the law of effect. *Abstr. Doctoral Dissertations, Ohio State Univer.*, 1953, No. 64.

Weimer, P. K. A proposed "automatic" teaching device. *IRE Trans. Educ.*, 1958, **E-1**, 51–53.

Zeaman, D. Skinner's theory of teaching machines. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art.* New York: John Wiley, 1959. Pp. 167–176.

# 39

# Educational Television — *Walter A. Wittich*

*Educational television is rapidly demonstrating its value in the educational enterprise. The author of the present article discusses the instructional role of television, its current educational programs, and its significance for teachers. Generalizations emerging from research are discussed and a table is presented summarizing the results of studies on the effectiveness of instruction by television. It would appear that the future holds an exciting and useful role for educational television.*

EDUCATIONAL TELEVISION officially made its appearance in 1952. Since the first channel for educational broadcast purposes was assigned to Houston, Texas, public schools ten years ago, 70 educational television stations have gone on the air and others are being planned. This growth parallels the rapid increase in general television use in the United States from fewer than 5,000,000 sets in use according to the 1950 census, to 65,000,000 sets in 51,000,000 household units, as shown in the 1960 census.

## The Instructional Role of TV

The mention of educational television can be counted upon to start a lively discussion. At one extreme are those who say educational television can do anything better. Those who are opposed minimize or deny its usefulness. The truth of the matter is being revealed by hundreds of research studies concerned with the classroom use of educational television.

Reprinted from *Educational Perspectives,* 1962, 1, 11–13, by permission of the publisher.

Since 1955, over 400 formal and informal investigations have been made of the role and the efficiency of educational television. These studies have investigated the use of television as a master teaching device, as a means of carrying large group instruction to hundreds or thousands of students, as a means of permitting the gifted learner to carry on independent study at his own pace and to the limits of his capacity to learn during sustained study periods. Investigations have measured educational television efficiency at the primary grade, high school, college, and adult levels, and in virtually all subject areas of the curriculum. Research results are summarized in Table 1.

TABLE 1. The Effectiveness of Instructional Television

| Level | | Math. | Science | Social Studies | Humanities History Lit., Arts | Language Skills | Health, Safety | Total | |
|---|---|---|---|---|---|---|---|---|---|
| 3rd–6th Grades | +TV | 14 | 8 | 12 | 0 | 14 | 2 | 50 | +TV |
| | n.s.d. | 21 | 14 | 11 | 0 | 36 | 4 | 86 | n.s.d. |
| | −TV | 3 | 1 | 1 | 0 | 10 | 1 | 16 | −TV |
| 7th–9th Grades | +TV | 4 | 9 | 0 | 2 | 0 | 3 | 18 | +TV |
| | n.s.d. | 11 | 8 | 1 | 7 | 0 | 1 | 28 | n.s.d. |
| | −TV | 2 | 3 | 0 | 0 | 0 | 0 | 5 | −TV |
| 10th–12th Grades | +TV | 0 | 3 | 3 | 4 | 1 | 1 | 12 | +TV |
| | n.s.d. | 10 | 7 | 17 | 17 | 6 | 0 | 57 | n.s.d. |
| | −TV | 5 | 3 | 0 | 9 | 4 | 0 | 21 | −TV |
| College | +TV | 0 | 1 | 1 | 0 | 0 | 1 | 3 | +TV |
| | n.s.d. | 4 | 26 | 24 | 11 | 12 | 7 | 84 | n.s.d. |
| | −TV | 0 | 1 | 4 | 3 | 1 | 4 | 13 | −TV |
| Total | +TV | 18 | 21 | 16 | 6 | 15 | 7 | 83 | +TV |
| | n.s.d. | 46 | 55 | 53 | 35 | 54 | 12 | 255 | n.s.d. |
| | −TV | 10 | 8 | 5 | 12 | 15 | 5 | 55 | −TV |
| | | | | | | | Total | 393 | |

Based on a recent analysis by Wilbur Schramm, "What We Know About Learning from Instructional Television," Institute for Communication Research, Stanford University, 1961. The table shows that out of a total of 393 TV research projects, 83 established television as a superior instructional technique, 255 reported no significant difference existing between television instruction and face-to-face or traditional instruction, and 55 reported a significant difference for traditional types of instruction.

This research supports several important generalizations:

1. Educational television makes it possible for large groups of students to learn as efficiently from television lessons as they can from face-to-face instruction. The comparison here is usually made between television instruction to groups as large as tens of thousands of pupils and the instructional outcomes which usually result when one teacher teaches a group of from 20 to 30 pupils. The implications of such large group instruction efficiency demand the study and attention of both professional educators and lay people.

2. Gifted pupils can learn more through television than through typical classroom instruction, and at accelerated rates, limited only by the individual's capacity and endurance.

3. Educational television makes it possible for a level of preplanning, rehearsal, and arrangement of teaching materials that is seldom possible in typical teaching situations. There are television teachers currently at work whose assignment is limited to one half-hour television lesson per day. However, a twelve to fourteen-hour day is barely enough time to permit the kind of endless planning, rehearsing, revising, and perfecting which typically precedes the best of educational television currently in use. With such concentrated effort, learning experiences of high quality can be provided through educational television.

The last five years have witnessed an exciting beginning of educational television research. No longer need it be argued that educational television can or cannot improve instruction. Today, rather, the basic problem is the manner in which subject areas, which lend themselves to television instruction, can best be served by television.

## Current Educational Television Programs

Under the direction of Superintendent Brish at Hagerstown, Maryland, one may witness one of the "oldest" research projects in this new medium. Begun six years ago, the Hagerstown television research was developed on a countywide basis in which outlying schools (some of them 40 miles away from Hagerstown) were interconnected by coaxial cables. At Hagerstown, a central television control center, surrounded by several rehearsal classrooms, went into operation with 22 television teachers involved in the demanding task of originating television instruction for use by classroom teachers. The television teachers at Hagerstown were chosen because of their interest in the medium and their background and ability in teaching. As each was confronted with the task of arranging for television teaching, he became involved in planning for the use of many audio-visual materials, as he sought to portray, through demonstration and the use of film clips, slides, models, charts, and graphics, the highest quality of learning experience. Needless to say, when one stands before one's colleagues, he is very much "on the spot," and what ordinarily would pass as good instruction in the day-to-day environment of a classroom no longer can be depended upon to meet the exacting standards of television instruction, which, among other things, attempts to compress a maximum amount of communication into a minimum amount of time.

Following years of planning, evaluating, and replanning, the Hagerstown television experiment now closely coordinates the planning of the classroom teachers with that of the television group. Representatives of the two groups have met in conferences to coordinate better the kind of teaching purposes envisioned by the classroom teacher with the unique strengths of the television teacher to supply learning experiences which ordinarily would lie beyond the time, budget, and mechanical opportunities of the classroom teacher to create. The classroom teacher does best what can be accomplished through discussion, interchange of ideas, question and answer, and reference and resource work, whereas the television teacher assumes the responsibility for the gathering together of fugitive visual materials which will enhance the learning experiences broadcast into the classroom. The result is an improved quality of day-to-day teaching which is possible only under such an arrangement: an "electronic" kind of team teaching is going on.

Another application of educational television is to be found in the program of the Midwest Program on Airborne Television Instruction, usually referred to as MPATI. After three years of preparation and with a budget of over $10,000,000 provided by business and foundations, including the Ford Foundation, MPATI began on September 11, 1961, a program of television broadcasts to an audience of an estimated 2,000,000 pupils and their teachers. By means of two television stations built into the fuselage of the aircraft, videotaped lessons are beamed to receiving sets in schools in six states.

Each day, Monday through Thursday throughout this school year, the "flying classroom" circles for five or six hours above the little town of Montpelier, Indiana. The airborne television impulses radiate more than 200 miles in all directions from a 24-foot sending antenna projecting vertically from the belly of the aircraft. The flying station broadcasts to schools in Illinois, Indiana, Kentucky, Michigan, Ohio, and Wisconsin. The MPATI curriculum includes a trial series of sixteen courses of instruction, ranging from elementary school through college subjects: music, foreign languages, arithmetic—including a special course for gifted children, science, and history. College students have classes in algebra, chemistry, and Russian. Eventually these taped courses may be available to Hawaii.

In broad terms, the purposes of MPATI are similar to those sought at Hagerstown, namely, to apply more careful planning to difficult portions of the curriculum, and thus to make possible learning experiences heretofore denied to the individual classroom. Through the screen of a television set, located in the independent classroom which may be hundreds of miles from the broadcast source, it is today possible to bring the advantages of modern communication tools into the most remote learning environment.

Another important kind of educational television involves the concept of total instruction. This refers to a process whereby an entire lesson is recorded on television film or video tape and becomes the major learning experience available to the student. This preplanned and recorded learning experience is supplemented by textbooks, study manuals, and an opportunity for feedback through means of interchange of written lessons. The key stimulus, however, is provided through recorded television teaching. When such learning is participated in by gifted students working individually or in small groups, remarkable outcomes have been recorded.

At the City Junior College in Chicago, it is currently possible to receive two years of junior college instruction entirely via television. The method of preparation of the learning materials is similar to those used in Hagerstown and MPATI. However, the student, seated as a remote classroom learner, is able to receive instruction through a television screen or in some cases, from a motion picture projector which projects the kinescope recording. He is thus able to proceed as rapidly as the broadcast schedule permits or as rapidly as he can witness and react to the film record of such broadcasts. In the City Junior College, 32 tests of television instruction have been undertaken in the fields of English, social studies, science, psychology, and the humanities. The data reveal that in 17 cases television instruction had as good results as face-to-face instruction and in eleven cases was superior. In two cases, classroom instruction was superior. Such research points to the possibility that large areas of traditionally taught subject content can be taught more efficiently through educational television.

A very special kind of television instruction is currently being broadcast from the College of Education of the University of Hawaii to teachers enrolled on Oahu and the neighbor islands. Through this 42-lesson televised course in methods of improving classroom instruction through the use of new communication media, a team of teachers is brought into the students' own local school room through television.

Thus Hawaii is involved in finding out whether or not television instruction can overcome the barriers of remoteness, distance, and accessibility to traditional teacher education facilities. The time can be envisioned when, in addition to residence instruction, large portions of the curriculum can be offered through television broadcast, through the relayed use of video tapes, or more important, through the independent learner's reacting privately to kinescopes of entire courses.

Instruction by television has been advanced markedly by the development of video tapes, a technique by which a program can be recorded on a wide ribbon of magnetic tape. The use of the kinescope makes it possible to transfer or to record a television program on a reel of 16mm sound motion picture film. The result can then be used much as a motion picture is used, either for classroom group instruction or for independent study. Other contemporary techniques include telefilm recording, video tape broadcast, and kinescope film broadcast. The latter two are coordinated with feedback techniques which involve individual long-distance telephone hookups, or signaling devices which are a part of the television broadcast network itself. In such ways the viewer can immediately record by electronic means his reactions to questions which flash instantaneously on the screen or to verbal questions by the television teacher. The results of such feedback can be made immediately available to the instructor to guide him in his explanations or to suggest a change of pace in his presentation.

## The Teacher and ETV

Teachers are likely to become deeply involved in educational television in the future. Therefore, some of the broad claims which can be made for television in classroom instruction or in professional education are here set forth.

1. Television is a remarkable, efficient technique for using cross-media communication tools. That is, television is the efficient carrier for all the modern communication devices—films, slides, tapes, opaque, vugraph, demonstrations, charts, specimens, models. As a master teacher or lecturer-demonstrator uses these techniques, television makes it possible for the individual viewer, even though he may be miles distant, to see the best of learning materials avaliable for educational use.

2. Student teachers of tomorrow will have the opportunity of viewing through closed-circuit or broadcast, the various aspects of pupil-teacher relationships which occur in the privacy of the classroom. All of a large group can see the same teacher with his class and thus have a common experience for later discussion and evaluation. At the same time, this use of television insures the privacy and the candidness of a learning situation which is usually altered when visitors walk into a classroom.

3. Educational television is a challenge to any teacher who is interested in seeing models of instructional technique which employ the so-called cross-media approach to learning. The best examples of educational television teaching employ a wide variety of modern teaching tools, e.g., 16mm sound motion picture films, slides, charts, tapes, models, specimens. Those who have had the opportunity of watching Professor Harvey White, as he has taught the continental classroom course in modern physics, or Dr. Peter H. Odegard as he conducted the recent NBC course entitled "The Structure and Functions of American Government," have seen outstanding teachers who employ, when they feel it desirable, the most appropriate learning tool during the learning process. Television thus can become a challenging experience for the teacher who would compare his own teaching techniques to those employed by television instructors who have at their command all of the resources of the modern communication world. Again and again, teachers who have observed educational television have reported that they were as interested in the teaching techniques as in the content presented.

The teacher is and must always be the judge of the usefulness of educational television as it can be used in his own classroom. Television must be judged, therefore, in much the same manner as are films, filmstrips, models, and other modern teaching tools.

Educational television, to paraphrase Charles Dickens, can be the best of things or it can be the worst of things. Television, when defined from the standpoint of its electronic characteristics, is but a carrier wave. As a carrier wave, it can communicate ideas; hence, what is carried is everything. Thus, it can be concluded that the output of television can be no more than its input.

What we put into educational television, for example, may be a demonstration of a teaching technique. If this is arranged by a poor teacher, who merely verbalizes before a background of gray drapes, the input which is caught by the television camera and broadcast is almost certain to be poor. However, a vastly different kind of output is possible when an excellent teacher goes through all the necessary stages to present an effective demonstration. The stages include careful preplanning, arranging for necessary materials, and rehearsing before a psychologist interested in laws of learning, an audio-visual communications expert, and an experienced educational television director.

*Effective* educational television, difficult as it may be to achieve, is surely the goal for all who work with this medium. The progress made in the ten-year history of educational television is enormous and yet the future seems unlimited.

# 40

# Laboratories for Foreign Language Teaching

## — *Milton L. Shane and Helen Lacy Shane*

*This descriptive article concerns another technological innovation within education—the language laboratory. Although many of the statements made about them are untested hypotheses, there appears to be much that can be gained through the use of language laboratories. Many of the principles underlying programed instruction also apply here. Still in a developmental stage, language laboratories offer many opportunities for research in the techniques of language teaching and in the results of such teaching.*

OUR EXPERIENCE with language laboratories has been at two levels: a college laboratory in operation since 1952, and a training school high school laboratory installed some nine months ago. Reflecting current trends, the laboratory is larger and much more expensive, and it provides for a somewhat greater number of students. The languages taught are French and Spanish.

Our approach to modern language teaching rests on several assumptions: (a) Vital national interests demand the training of increasing numbers of Americans to be proficient in one or more modern languages. (b) Proficiency in modern languages will generally be used to increase professional or technical efficiency in nearly every field conceivable. At the same time the demand for trained modern language teachers is rising sharply. (c) Under

Reprinted from *Educational Leadership*, February, 1961, 18, 293–298, by permission of the senior author and the Association for Supervision and Curriculum Development.

these conditions, proficiency means first of all the ability to *understand* a foreign language when spoken by natives, and to *speak* the foreign language so as to be readily understood by natives. The skills of reading and writing follow and depend upon the skills of listening and speaking. (d) The acquisition of the listening and speaking skills is based upon neuromuscular memorization (analogous to learning to play a musical instrument), and must involve prolonged and regular periods of drill.[1]

These principles were successfully applied in ASTP language programs as early as 1943, but such programs were too costly and too cumbersome for nonmilitary institutions. The modern language laboratory has been devised to provide something like the ASTP learning experience in a more practical form. The laboratory does provide the voices of native speakers as models for learners, and it makes possible effective simultaneous practice in large groups.

Let it be said at once that this implies full-time use of the modern language laboratory. Every language class must devote some time to laboratory drill every day. The modern language teacher cannot afford to share the laboratory with teachers of other subjects, any more than a music teacher can afford to share her instruments, or a typing teacher his typewriters.

Although we at Peabody have made no attempt to use the laboratory in teaching foreign languages in the elementary school, we do not believe that laboratory instruction is suitable for the preadolescent child.

## Laboratory Equipment

For teachers who wish to develop listening and speaking proficiency in their students, a record player and a tape recorder are the minimum essentials. Using suitable recorded materials, these machines make possible much practice in listening and repeating, in concert and individually.

The fully equipped laboratory adds to the experience of listening and repeating, the means of self-criticism by recording and playing back. Each student works in a semi-private booth, equipped with headphones, microphone, and a tape recorder. He listens to prerecorded material, spoken at normal speed by native voices, repeats what he hears in pauses left for the purpose, and then is ready to play back the recording of his own repetition in alternation with the native model. Unlike any other type of language learning activity, this permits every student to recite simultaneously and receive correction in the laboratory period.

---

[1] A more detailed statement of these principles will be found in: Johnston, Remer & Sievers, *Modern foreign languages*. U.S. Office of Education Bulletin 1960, No. 20. Washington, D.C.: Superintendent of Documents, U.S. Government Printing Office. 30 cents.

The Peabody College laboratory is of the library type. Students in first and second year classes are required to schedule two laboratory periods per week, in addition to four periods of class attendance. In the laboratory each student checks out a tape, goes to a booth, and works independently of others, listening, repeating, recording, and playing back such material as he thinks most needed, throughout the period. The laboratory attendant checks the roll, issues tapes, and gives assistance in operating equipment if needed. Another assistant works week-ends, duplicating tapes as needed. The services of an electronics repairman are available.

This type of laboratory, providing maximum opportunity for independent preparation, seems satisfactory at the college level. We are convinced that it would not work well with high school students.

An alternative form of laboratory is the console or broadcast type, in which the instructor controls the rhythm of listening, repeating, recording, and playing back through a central console; and at the same time monitors the work of individual students in turn, using a two-way communication system to give corrections as needed in absolute privacy.

The fully equipped console laboratory provides a tape recorder for each student. The tape recorder may be in the student's booth, operated by the student; or miniaturized tape recorders may be stored in a remote control cabinet, and operated through the central console. This latter plan obviates tape spillage, unintentional tape erasure, emergency splicing, and various types of mechanical mishap incidental to student operation of tape recorders. The tape recorders in the remote control cabinet are loaded with endless-loop cartridges, which run forward only and provide for playback at the end of a cycle—usually three or five minutes.

The system just described (the Linguatrainer) was originally developed for the Massachusetts Institute of Technology. It is the one installed at the Peabody Demonstration School. We find it well adapted for use during class periods. Each class spends a part of each period working in the laboratory, and the rest in conventional classroom activities.

Indispensable accessories to this (or any) language laboratory are a tape recorder, a record player, and a library of tape and disc recordings. These are used to transfer prerecorded material to the cartridges used in the Linguatrainer system; or to the library tapes used by the students in the college laboratory.

Indispensable, too, are the services of an electronics repairman who has become familiar with the laboratory and is readily available. Emergencies will arise. It is important to have spare parts always on hand. A wiring diagram should be on file, particularly if a substitute repairman has to be called in.

## Materials and Methods

As previously implied, proficiency is achieved by listening to recordings of native speech and imitating, repeating, correcting, and repeating until new speech habits are firmly established. Materials to be memorized consist of dialogues and pattern drills. Dialogues, which are the core of each lesson, present real-life situations of wide interest and are arranged in a suitable sequence for learning structural forms of the language. Pattern drills are suggested by passages in the dialogues, but give additional drill on new structural patterns as these are added.

Teachers who meet four or five classes per day will not have time to create new textbooks of this type. Fortunately, a number of college texts in French, German, Spanish, and Russian (inspired by the ASTP Spoken Language manuals) have come on the market since 1950, and are now available completely tape-recorded. A few high school texts have recently been modified for laboratory use, and complete tape recordings of these texts are available. We have one completely new textbook, written to meet the needs of present-day teaching organized around the language laboratory. This is *Modern Spanish* (published by Harcourt, Brace and Company, 1960), written by a committee of authors as a project of the Modern Language Association. It is tape-recorded. So far as we know, no comparable completely new text has been produced at the high school level for any modern language. Such textbooks are badly needed.

Theoretically, any teacher who speaks the language natively can prepare his own recordings; but this is an enormous task. It also limits the students to hearing one voice, rather than the variety of men's and women's voices now available in all the commercially taped texts referred to previously.

It should be clear at this point that teaching techniques figure largely in the selection of a textbook, and in the day-to-day preparation (or re-taping) of taped material for laboratory use. Once the student is in the laboratory, the nature and organization of the material will largely determine and control activities.

Note, however, that the laboratory makes it possible for retarded students to work at earlier lessons, or specially assigned material, without hindering the progress of the group. Exceptionally able students may likewise work with special materials.

Additional laboratory periods may be scheduled for those who need or wish them. In a high school situation, students may spend study or library periods in the laboratory when this is thought advisable. Through the console, such pupils may be assigned individual tapes with which they will work in a library situation. This assumes that the laboratory has a few booths available each period, over and above the number in the largest class.

At the second-year and more advanced levels, relatively few taped textbooks are as yet available for laboratory use. There is, however, a wealth of literary material recorded on disc and tape. The ingenuity of the teacher will find full play in adapting such material for laboratory use: inserting pauses for repetition, adding drills to fix new details of structure, providing questions based on the text, spacing passages to be used for self-dictation, and so forth.

Similar use may be made of foreign language programs taken off the air, particularly newscasts.

## Testing

As the laboratory is used to develop the skills of listening and speaking, it is these skills which can be tested there to advantage. Actually, group tests of listening comprehension, such as those published by the Educational Testing Service, need not be given in the laboratory. Students in a conventional classroom may listen to a tape recording and mark multiple-choice responses on an answer sheet. However, acoustical conditions are unequal unless the test is given in the laboratory, where every student hears the recorded material with equal volume and fidelity. Shorter listening comprehension tests of this type may be prepared by the teacher and given in the laboratory.

The most difficult skill to measure is that of speaking. Heretofore this was tested through interviews, which were time consuming, and difficult to score objectively. Laboratory tests of speaking reduce the testing time 50 percent, since an entire group may record responses simultaneously. The simplest type of test is one in which the student is asked to repeat sentences (arranged in ascending order of difficulty) as he hears them. Scoring is done on predetermined points of pronunciation and intonation, as well as memory span. Next come tests in which the student is asked to modify sentences in accordance with a model which is given. These are simply pattern drills with the correct response omitted. Such tests may range from very simple to very difficult. Finally, students may be directed to ask or to answer questions.

It goes without saying that no exercise should be given as a test until students have practiced thoroughly the equivalent drill. Moreover, frequent practice tests on familiar material will reduce the tension created by this unfamiliar kind of examination.

It will be noted that tests consisting of responses recorded by students in the laboratory, must be heard and scored in the laboratory. In the console laboratory, the instructor can monitor all student tapes through the console, one after the other, when the group examined has vacated the laboratory; but this is a lengthy process, and the work must be completed before another group uses the laboratory, unless it is possible to withdraw the cartridges or tapes carrying test material, shelve them, and insert fresh tapes for the next group of testees.

A much cruder technique is to test students in the laboratory one at a time, but briefly. In this instance each student in turn hears the same test recording and each in turn records his responses on a single large reel of tape. This tape, carrying the test recordings of an entire class, may then be played and scored anywhere at the convenience of the instructor. However, this plan creates more nervousness on the part of testees, since the familiar laboratory situation is somewhat modified.[2]

One item of equipment which may be on the market, though we have not seen it, is a portable record-playback unit for tape cartridges. This would be very helpful in the preparation of materials. It would likewise be useful for scoring student recordings away from the laboratory.

Speaking in general, we hope and are inclined to believe that equipment will soon be devised to make the testing process less cumbersome than that we have described here.

Giving free rein to the imagination, may we suggest the invention of an automatic critic, a device which will scan student recordings while they are being made, and reject speech which is too far below the norm of a given level. In the classroom this is done by the teacher, though necessarily at random. In the small sections of the ASTP it was done much better by the native informant. In the laboratory, the alert student practices self-criticism as he compares his recorded repetition with the master tape, phrase by phrase. Not all students are alert, however, and the beginner sometimes fails to detect significant differences. This is a real challenge, which we believe the inventor will ultimately meet.

Returning to the level of what is possible now, we must stress again the need for a much greater variety of taped teaching material, especially at levels beyond the first year.

Something very helpful to teachers who are "playing by ear" in the laboratory would be a sound film, showing in a good deal of detail the operation of one or several long-established laboratories, at the high school and the college level.

The film just mentioned suggests the importance of in-service training. For a long time to come, teachers who have had some experience but have been feeling their way in the use of the laboratory will welcome the opportunity to see techniques demonstrated, and to compare notes with other laboratory teachers.

---

[2] For more detail on pattern drills and audio-lingual testing, see Edward M. Stack, *The language laboratory and modern language teaching*, London: Oxford University Press, 1960, and Pierre Delattre, Testing students' progress in the language, in *Language teaching today*, Indiana University Research Center, October 1960.

Today, a steadily increasing group of teachers with no previous experience is being asked to use the language laboratory. For them, in-service courses will provide an opportunity to practice the mechanics of laboratory operation away from the tensions of the classroom. At the same time (and this is most important), language teachers of many years' experience in conventional teaching may obtain needed help in adjusting to the new objectives which the laboratory was created to attain, and in using the materials now becoming available for laboratory teaching.

# 41

# Team Teaching: An Assessment

## — *Harold D. Drummond*

*Team teaching is one among several new administrative organizations for instruction. Citing the lack of precision in educational terminology, the author describes five types of team teaching in various stages of development and/or experimentation. He evaluates "nine advantages" consistently reported for team teaching in terms of an explicit set of assumptions, and calls for more extensive research and experimentation.*

INTEREST IN team teaching has grown rapidly since 1958. In that year the National Association of Secondary-School Principals devoted the January issue of *The Bulletin* to the work of its Commission on the Experimental Study of the Staff in the Secondary School (NASSP, 1958). In the two-year period preceding that issue (1955–57), not a single article on the subject of team teaching was listed in the *Education Index*. In the following two-year period, by contrast, the *Index* listed eight articles. Between July 1959 and June 1960, 35 articles appeared in published journals and, through June 1961, an additional 19 entries had been listed. No doubt the output will continue to mount, for everyone seems to want to jump on the bandwagon of "team teaching" (Bush).

At the moment, it appears likely that in hundreds of secondary schools and in many elementary schools the instructional staffs are doing something

Reprinted from *Educational Leadership*, December, 1961, 19, 160–165, by permission of the author and the Association for Supervision and Curriculum Development.

which they call team teaching. What types of team teaching are reported by school systems? What are characteristics of present developments? What advantages are claimed for team teaching, and what problems are inherent in the structures already adopted?

## Types of Team Teaching

The education profession has suffered for years because it has lacked precise terminology. Team teaching is another example—the term already has almost as many meanings as there are school systems doing something with it. At present, there appear to be at least the following five types of team teaching in various stages of development and/or experimentation. Variations from these types are, of course, myriad.

1. *A hierarchy of teaching assignments.* Several school systems (see Anderson, Johnson, Stone) have attempted to develop instructional teams which are based upon a specified hierarchy of teaching assignments. At the top of the hierarchy is a team leader who is a person with superior educational preparation, several years of teaching experience, and leadership qualities. The team leader often is given a lighter teaching load and a salary commensurate with the leadership responsibilities he is asked to assume. The team, in school systems developing hierarchal assignments, usually also consists of senior teachers (who receive extra pay, but not as much as that received by the team leader), regular teachers (often those without previous experience or those new to the system), part-time teaching assistants, and clerical aides. In order to cover the costs of the increased salaries for leadership and for clerical help, additional pupils are assigned to the team —usually at least one extra class section for three or four certified teachers.

2. *Coordinate- or co-teaching.* In school systems using this approach, teachers are assigned to a large group of pupils (usually a multiple of the number the teachers would have under more traditional assignments; e.g., two teachers to 60 youngsters, three teachers to 90) and they plan together *as peers* how best to provide for the pupils for whom they are responsible. As in the previously described "hierarchy" plan, sometimes instruction is provided to the entire group by one teacher. Sometimes one teacher works with most of the youngsters in the group while the other works with a small group of the gifted or with those needing remedial instruction. Sometimes each of the teachers has a "normal-sized" group of about 30 pupils each. Attempts are made in the planning to utilize to the fullest extent the strengths of each teacher. Such plans usually have been described as existing within established departments at secondary school levels or at grade levels in elementary schools.

3. *Team teaching across departmental lines.* In several junior and senior high schools attempts have been made to improve the program, and hopefully to improve learning, by devising schedules for instructional teams which provide a two- or three-period block of related content (e.g., American history, American literature). Students have, normally, one period with the social studies teacher, followed by one period with the English teacher (or the reverse). Often, when desirable, the two groups are combined for the double period—as for a field trip, orientation to a new unit, lecture by an outstanding resource person, visual aids, and the like. The teachers have at least one free period at the same hour so that joint planning is possible.

4. *Part- or full-time helpers.* Many descriptions of team teaching indicate a fairly standard teaching role for the regularly certificated teacher, but seek to improve his teaching effectiveness by providing additional help of various kinds, including instructional secretaries, theme or paper correctors, laboratory assistants, learning materials coordinators, and audio-visual experts. To employ the additional personnel without substantial increases in instructional costs, teachers usually are asked to accept responsibility for a larger number of learners than normal (usually 35 to 40). The teacher retains active control of the planning and most instructional phases of teaching, utilizing the helpers on the team for particular tasks of a more routine nature.

5. *Trading groups.* In an informal way this method of capitalizing on the particular strengths of teachers has been utilized for years by elementary school teachers. The teachers have said, in essence, "If you'll take my art—you're good in it and I'm not—I'll take your music," or "If you'll take my science, I'll take your social studies." Until recently, such "trading" was rare at the secondary school level, but it may be growing now as a result of the staff utilization studies.

Several reports indicate that two or three teachers of a particular subject, such as general science, plan their work so that they trade groups for particular units of content. The trading is done, ostensibly, to make certain that the groups receive instruction from the best-qualified teacher of the team, and also to ensure that the teachers have an opportunity to provide instruction geared to their own interests and competencies.

## An Assessment

Any attempt at assessment of educational practices is, of necessity, made from a value base. In most previously published reports, an attempt has been made to assess practices in team teaching by utilizing three types of data: achievement as measured by standardized or by locally-constructed tests, teacher opinions (sometimes buttressed by student and parent opinions), and per-pupil costs. The data collected and reported generally indicate: (a) Students do as well or perhaps a little better on standardized tests when taught by teaching teams of the various types described. Usually the obtained differences are not significant when fairly sophisticated statistical measures are employed to analyze the data. (b) Teachers, generally, are willing to continue the team approach, although there are numerous indications that not all teachers make good team members. Increasingly, reports indicate that differences among teachers need to be recognized equally as much as do variations among learners (see Hanvey and Tenenberg, Weiss). The reports seem to show a feeling of, "We are working on the frontier—trying to find a better way of proceeding," which undoubtedly has positive value for heightened morale. The increased workloads (meetings, meetings, meetings!) seem to have been shouldered with enthusiasm by the participants. In the long pull, better ways of equalizing instructional loads probably will need to be developed or morale may slip. (c) Students and their parents generally favor what has been tried. Many learners are at first skeptical or negative, but as teachers gain confidence and competence in their changed roles, reports from them indicate positive support for the team approach. (d) Costs rise slightly. The extent of the increased costs usually is not specifically reported. Three ways of bearing the increased costs have been utilized: increased local appropriations, employing fewer qualified teachers and increasing the pupil-teacher ratio, and support from foundations. Many of the additional costs have been the result of improved instructional resources—books, films, overhead projectors, and the like.

While these criteria of achievement, opinion, and cost are measurable, to some extent at least, they do not necessarily provide good bases for assessment unless one subscribes to the following premises: (a) that education is best which results in highest achievement as measured by tests, standardized or otherwise; (b) that education is best which results in expressed teacher satisfaction with administrative practices (and perhaps student and parent approval also); (c) that education is best which increases present per-pupil costs only slightly and may in time tend to lower costs. These premises seem to be questionable as criteria for a profession to use in assessing the worth of an innovation.

The assessment which follows, also made from a value base, is developed to the extent possible on the following assumptions: (a) learning of high quality requires interaction between the teacher and the learner and between the learner and other learners; (b) learning of high quality is more likely to occur when teachers are patient, understanding, intellectually alert, and free to make decisions based upon their best professional judgment; (c) what is learned must be used (more functionally than on an examination) or before long it will not be known.

These assumptions obviously eliminate cost as a function of quality (although most administrators at present must consider the cost-quality factor) because the writer assumes that this nation *can* afford instruction of high quality for its children and youth. The assumptions also eliminate teacher opinions as expressed on questionnaires or verbally to members of the administrative and supervisory staffs. What is essential for effective learning is not necessarily highly correlated with what teachers prefer. To state the assumption another way, what teachers consider to be good teaching may not result in the most effective learning. The spotlight should be focused on the *learning process* rather than on teaching.

These assumptions also eliminate achievement as measured by tests. Teachers know what most achievement tests contain or are likely to contain. Using almost any organizational structure, they can, therefore, make sure that the learners make about average gains in achievement. Obviously, any structure which results in *marked* improvement on standardized tests should be seriously considered. Whether the instructional technique or structure should be adopted widely, even if better test results are obtained, is another matter—a matter for professional judgment. Students who score higher on standardized tests, in other words, are not necessarily better educated.

What assessment can be made of the various types of team teaching using the value assumption that good learning results from the interaction of learner and learners with patient, understanding, intellectually alert, free teachers who see that what is learned is used? Nine "advantages" consistently reported for team teaching are given in italics below. In each instance, some comments based upon the value orientation of the writer follow in regular type.

*Few pupils are limited to the instructional competence of a single teacher at a grade level or in a department at the secondary level.* As a result, few teachers in this arrangement get to know individual pupils as well as in traditional arrangements. Interaction between the superior teachers and the learners (especially in the hierarchal plan) is minimal. Personal contacts of learners with teachers tend to be limited to teachers of lesser competence and experience.

*Persons most highly qualified provide instruction to large groups, thus saving much time for the total staff which can be used for more effective planning and for instruction in smaller-than-average groups.* Questions learners have during the lecture must be deferred until a later time. Moreover, what the teacher wants to teach is not necessarily what the pupil needs or wants to learn. The learner may, in fact, already know what is being presented to a large group. The same problem exists, of course, when teachers lecture to normal-sized groups. May there not be a better way to teach?

*In presentations to large groups, better use is made of visual aids because more time can be devoted to the preparation of needed materials by specially-qualified team members.* Subsituting a picture of a magnet on an overhead projector as a lecturer explains how it works (as was shown recently in the television report, "The Influential Americans") may result in undesirable verbalization not sufficiently based on real, firsthand experiences by the learners themselves. Skillful presentations do not necessarily result in effective learning experiences.

*More uniformity in instruction is achieved because all students are taught, both in the large groups and the small, by the same teachers. Sections pupils are assigned to thus make less difference than in traditionally organized schools.* Uniformity in instruction is not necessarily desirable. The degree of desirability depends largely upon how much flexibility is provided for the very bright students and the slow learners. Individualization of instruction, whether in traditional or team approaches to teaching, is a valid and desirable goal. To the degree that attention to individual differences is provided (this varies in different team teaching plans), the learning is likely to be effective.

*Less repetition is required of teachers, especially at the secondary level where several sections of the same class have been traditionally assigned.* Repeating a lecture to several sections of the same class probably *is* wasteful—but getting to know the pupil is *essential* for interaction. Almost all reports indicate that less discussion occurs as team teaching is undertaken. Perhaps more "ground" can be covered, but that is no guarantee that more learning has taken place.

*Teacher competencies are better utilized.* Instruction tends to become more formal, less spontaneous. In the hierarchal plan, young, inexperienced teachers undoubtedly have more opportunity to learn from team leaders, but the conception is supported that superior teachers lecture to large groups while teachers drawing lower salaries and with less teaching experience work with smaller groups. Learners, as a result, get individual help from teachers who probably are least qualified to give it. These weaknesses, it should be noted, are not apparent in the coordinate and interdepartmental plans where teachers operate as peers.

*Better provisions are made for helpers—librarians, audio-visual experts, clerks, and the like—to do routine tasks.* A definite boon to the profession! The only problem which should be noted: effective coordination of such helpers takes time. In the opinion of this analyst, such help should and could be provided regardless of the structure for teaching developed by the school system.

*Group size is clearly related to function. Large groups are formed for activities which are effective with large groups, and vice versa.* This concept makes sense. In the judgment of this assessor, the "coordinate" and the "across departmental lines" teams have the greatest possibility of built-in flexibility at this point. The "hierarchal plan," because of the specified roles, probably has the least likelihood of achieving flexibility in grouping.

*Of necessity, students assume more responsibility for their own learning. As more and more instruction is provided in large groups, a greater share of the school day is given to independent study on the part of learners.* If education is effective, the more mature the learner, the more able to guide and direct his own learning endeavors he should be. Generally, then, this claimed advantage of team teaching is desirable. Perhaps even a greater measure of independence could be achieved other ways, however.

## A Final Word

The worth of attempts at team teaching are not proven to date. The main value of the attempts which have been made thus far undoubtedly lies in the staff growth which has occurred as a result of the experimentation.

Experimentation should be continued. Much more sophisticated research designs should be used, so that variables in the situations can be more carefully controlled. While team teaching is being tested more carefully, some school systems (perhaps the same ones) should also be testing other approaches to improvement of learning, such as: assigning not more than 20 pupils to a teacher, shortening the teacher-directed part of the school day and lengthening the pupil-directed portions of the day, utilizing more programmed materials as these become available, basing more instruction on the "workshop way of learning," orienting in-service education programs for teachers more toward intellectual growth, providing better learning materials centers and instructional secretaries in every school, and lengthening the school year for a larger number of teachers so that more time for planning and preparation is available.

## BIBLIOGRAPHY

Anderson, R. H. Team teaching in action. *Nation's Schools,* May 1960, 65, 62–65.
Anderson, R. H. Team teaching. *NEA J.,* March 1961, 50, 52–54.
Anderson, R. H., Hagstrom, E. A., & Robinson, W. M. Team teaching in elementary school. *Sch. Rev.,* Spring 1960, **68** (1), 71–84.

Bruntz, G. G. Team approach to social science teaching. *High sch. J.*, April 1960, 43, 370–374.

Bush, R. N. Editorial: The team teaching bandwagon. *Calif. J. secondary Educ.*, April 1960, 35, 207–208. (This issue and the February 1960 issue of the same publication contain a number of articles on team teaching most of which are not listed individually because of space.)

Ford, P. M. Different day for the English teacher. *Eng. J.*, May 1961, **50**, 334–337.

Hanvey, R., & Tenenberg, M. S. University of Chicago Laboratory School, Chicago, Illinois, evaluates team teaching. *The Bull.*, National Association of Secondary School Principals, January 1961, 45, No. 261, 189–197.

Johnson, R. H., & Lobb, M. D. The transformation of the sacred secondary school schedule. *Calif. J. secondary Educ.*, February 1960, 35, 96–105.

Mahoney, W. M. Try coordinate teaching. *Amer. sch. board J.*, November 1959, **139**, 13–14.

National Association of Secondary-School Principals. *The Bull.* Complete issues devoted to staff utilization studies include: January 1958, 42, No. 234; January 1959, 43, No. 243; January 1960, 44, No. 252; January 1961, 45, No. 261. (These bulletins comprise the most extensive reporting available—each issue containing reports from systems undertaking experimentation under the Commission on the Experimental Study of the Utilization of the Staff in the Secondary School. Most individual reports are not listed herein, to conserve space.)

Smith, V. H. Team teaching has its advantages. *Eng. J.*, April 1960, **49**, 242–244.

Stetson, G. A., & Harrison, J. P. Junior high school designed for team teaching. *Amer. sch. board J.*, May 1960, **140**, 38–42.

Stone, W. J. New designs for secondary school scheduling. *Calif. J. secondary Educ.*, February 1960, 35, 126–130.

Ward, J. O. Another plan for coordinate teaching. *Amer. sch. board J.*, February 1960, **140**, 10.

Weiss, T. M., & Morris, Mary Scott. Critique of the team approach. *Educ. Forum*, January 1960, 24, 207–208.

# 42

# Nongraded Programs: What Progress

## — *Hugh V. Perkins*

*The author indicates that nongraded programs have been used primarily as organizational devices in grouping pupils, but sees them as potentially becoming much more than this. Advantages and disadvantages of nongraded programs are discussed and the need for further research is noted.*

AMONG THE most perplexing realities which continuously recur in the life of every teacher are the marked differences in achievement, intelligence, motivation, cultural and experiential background, and physical and social maturity which exist within any group of learners. For decades educators have given lip service to the need for teaching children as individuals and for providing more adequately for individual differences. They have conducted extensive research in child development over nearly a half-century, have studied insights gained by teachers during two decades of participation in child study programs, and have noted the continuing research in perception, learning and personality. All these have demonstrated that learning is most effective when it is meaningful and is related to the individual needs, perceptions and interests of the learner, when it begins where the learner is, and when it is perceived by the learner as enhancing his own self-concept.

Yet, in spite of these influences, until recently no break-through has been achieved toward fully implementing the principles of individual differences.

Reprinted from *Educational Leadership*, December, 1961, 19, 166–169, by permission of the author and the Association for Supervision and Curriculum Development.

The critical reexamination of American education which has been carried on by educators and lay groups since 1957 has resulted in the development of fresh new approaches to age-old problems in education. During this period also, novel but promising approaches toward solution of an educational problem which heretofore has gone largely unnnoticed, have suddenly come under closer scrutiny and have become the focal point for intensive study and experimentation.

One such approach and type of organization is the nongraded school. This arrangement provides a flexible grouping and organization of curriculum content that enable children of varying abilities and rates of maturing to experience continuous progress in learning. Although the specific details of nongraded organization vary in different school systems where it is being used, certain general characteristics may be noted.[1] In most of these schools the nongraded organization functions only during the primary years, but often is linked with the kindergarten program. A few schools are experimenting with this type of organization in the intermediate years.

The nongraded type of organization appears to have had the greatest influence on the program and pattern of grouping in reading. Characteristically, the reading experiences of the primary years are organized into eight to ten sequential steps or levels through which each child moves at his own rate. Probably its most unique feature is the flexibility possible in moving a child from group to group or from one classroom to another at any time his growth warrants such a change.

In an effort to assess the present status of nongraded programs, this writer analyzed questionnaire responses from 14 representative systems, many of which have used some type of nongraded pattern of organization for three years or more.[2]

## Impetus for Initiating Program

The most frequently mentioned reason for adopting a nongraded type of organization was a growing dissatisfaction which teachers and administrators expressed regarding promotional policies they had been using. In many cases this dissatisfaction was echoed also by parents. Several respondents felt that the nongraded program might give them a longer period of time for studying the progress of individual pupils before making a decision on retention. It was hoped, too, that such a plan might result in larger numbers of children completing the primary program in the normal period of time. Many respondents mentioned their desire to eliminate the feeling of frustration children have as the result of failure.

---

[1] For a comprehensive discussion of the nongraded school, see John I. Goodland and Robert H. Anderson, *The Nongraded Elementary School,* New York: Harcourt, Brace and Co., 1959. Also see various articles appearing in educational journals.

[2] The writer is indebted to the following school systems which responded to this survey: Corona and Torrance, California; Pocatello, Idaho; Moline and Park Forest, Illinois; Baltimore and Germantown, Maryland; Marblehead, Massachusetts; Dearborn, Michigan; Reno, Nevada; Dayton and Youngstown, Ohio; Savannah, Georgia; and Milwaukee, Wisconsin.

An important impetus in the initiation of these programs was the strong conviction held by key administrators and teachers that educational practice should be more consistent with what is known concerning individual differences. The nongraded types of organization seemed to hold promise of fuller realization of development and learning for all children. Other educators were dissatisfied with the achievement of their students in reading and language arts. They believed that reading performance would improve in a nongraded program in which children move to successive levels only when they are ready for that level.

What were the educational and philosophical considerations which prompted these respondents to change their attitude toward a nongraded program?

The administrators and faculty members believed that: (a) learning should be continuous; (b) children grow and learn at different rates and each should have the opportunity to achieve at his own rate; (c) the school program should be flexible so as to meet varying developmental needs and growth patterns of individual children; and (d) greater achievement will result when children experience success in school.

Nearly all respondents indicated that an initial and continuing orientation is used to acquaint teaching staff and parents with the objectives and operation of the nongraded program. Typically, a study by the faculty has preceded its introduction. In some schools parents have participated with teachers in the preparation of orientation materials and in the development of a new report card. It seems clear that the general acceptance of the nongraded program by communities using it is due in no small measure to the careful and continuing orientation of both staff and parents.

Respondents report that the chief advantages of the nongraded pattern of organization are these: fewer children are retained, children's progress in learning is greater and more continuous, flexibility in grouping has succeeded in fitting the educational program more closely to the needs and maturity of the individual child, and pressures for achievement and maintaining standards have been eliminated or markedly reduced. In addition, some nongraded schools report: an improvement in teacher-parent rapport; slower children are identified earlier; higher teacher and pupil morale; and increased teamwork among teachers and between teachers and parents. Moreover, when learning experiences are tailored to the needs of the individual, curriculum takes on new meaning. The flexibility inherent in the nongraded program permits the use of team teaching approaches and allows teachers to work with the same group more than one year.

Disadvantages noted were that some teachers and parents find absence of specific grade standards difficult to accept. Teaching a nongraded group requires more work of the teacher, but it also is apparently more satisfying. Several respondents mentioned the increased time and effort required in keeping records. Another mentioned the problem of articulating the "levels" program with other subject areas such as arithmetic and social studies. Finally, several report problems relating to the orientation and adjustment of new teachers to the nongraded program. One reports difficulty in finding teachers who are willing to work with the less mature or the slow learner. Another holds that a few teachers may use absence of grade levels as an excuse for their own indifferent performance.

## More Evaluation Needed

Much of the evaluation of nongraded programs has been informal and has consisted of questionnaire reactions of parents and teachers, observed behavioral changes in individual children, and comparisons of achievement test scores. Collectively, these evaluations show that children are making significantly greater progress in reading achievement than are matched children in traditional classrooms, and that the reactions of parents and teachers are in general favorable.[3] The nongraded pattern appears to be especially facilitative in enabling the more capable children to achieve advanced reading skills and to proceed toward independent reading much sooner than previously.

Shapski found that children in a nongraded program scored significantly higher than did matched controls in traditional graded schools. She also found that children at all ability levels benefited from the individualized flexible program, with those of superior intelligence making the greatest gains of all. In addition, Shapski found that under the nongraded plan less than half as many children of average or low average intelligence spend four years in the primary program as would occur if the decision on promotion was rendered at the end of their first year of school.[4]

The marked increase in interest shown and the adoption of some kind of nongraded organization by hundreds of schools each year indicate clearly the need for more controlled and definitive studies of nongraded programs. What are the effects of this type of experience on other areas of children's learning and development? We need to learn much more concerning the effects of these programs through carefully controlled studies which assess the changes in children's self concepts, levels of anxiety, levels of aspiration and feelings of self-esteem, increase in peer status, and changes in other skill and subject areas. Only then can the effectiveness of these programs be ascertained, the findings used to make needed improvements, and the knowledge gained for adapting and modifying a general structure to fit the specific needs of a particular school and community. Some of these studies are already under way, but many more are needed.

---

[3] Vivien Ingram, Flint evaluates its primary cycle, *Elementary Sch. J.*, November 1960, 81, 76–80.

[4] Mary K. Shapski, Ungraded primary reading programs: An objective evaluation, *Elementary Sch. J.*, October 1960, **61**, 41–45.

## A Look to the Future

The mounting widespread interest in and acceptance of the nongraded program mark this movement as one which is likely to have a strong influence on future educational programs and organization. It is evident that this arrangement implements in educational practice much that we now know about individual differences. Too, in the freedom given the learner to proceed at his own pace, in the lessening of pressures, and in opportuinties given each child to experience success, the nongraded program brings to fuller realization the qualities of a favorable climate for learning.

On the other hand, some rather obvious limitations and pitfalls have been noted in nongraded patterns of organization. First, in most nongraded schools the individualization of instruction has been confined to reading. While reading is a skill of central importance, it is apparent that individualization and pacing of instruction leading to increased achievement in other areas also are long overdue. Secondly, there is danger that without a strong commitment to a program based on the individual rates of maturing and needs of children, the sequence of stepwise levels may result in the replacing of grade standards by another set of standards different in name only.

Finally, with the nongraded pattern having proved successful in many schools which planned carefully in orienting teachers and parents prior to its introduction, there is great danger that other school systems will jump on a nongraded bandwagon without an adequate preparation of staff and community. Experience suggests that schools should prepare carefully and begin slowly so that the nongraded program which emerges is one that is tailored to particular children of a particular school and community. Planning for children's development should not result in a mad dash for the new look merely so that educators and parents may feel that they are educationally outfitted in the latest style!

Much of the potential of the nongraded patterns is as yet untapped.

Thus far, it has been used primarily as an organizational device for grouping pupils, yet it has the potential for becoming much more than this. This more valid view of children and their development calls for new approaches to the curriculum. The exploding increase in knowledge emphasizes the need for sequential organization of content so that individualization and pacing of instruction can be carried out in all fields. In truth, our lives are far too short and our needs are too great for us to afford the luxury of children's being bound to rigid standards and inflexible segments of content organized as grades through which students pass in lock-step fashion.

The nongraded pattern of organization, together with the body of philosophical and psychological principles which give it meaning, has the opportunity for influencing profoundly the pattern and organization of elementary education in America and possibly secondary and higher education as well. The extent to which this influence results in educational change will depend upon those of us who use and develop its pattern and philosophy.

A look into the future might hopefully reveal innumerable varieties and types of individualized, self-paced patterns of instruction, each developed as a result of thoughtful planning and continuous evaluation in relation to the unique needs of students and community. In a broader context, one might hope that the knowledge gained in the development of this and other instructional patterns might lead to a more comprehensive and functional understanding of how human beings learn and develop. The extent to which these hopes are realized depends upon the wisdom and commitment of those of us who labor in these vineyards.

# 43

# Administrative Leadership — Key to Core Program Development — *Gordon F. Vars*

*Assuming that block-time, unified studies, and core programs—when carefully developed and supported—can provide effective means to attain the goals of education, the author draws upon extensive survey material in discussing the factors that should be considered in establishing and maintaining such programs. Emphasized is the necessity for continuous support—both material and psychological—from administrators. The potential advantages of these programs in terms of academic and nonacademic outcomes seem to be worth the administrative reorganization and effort that would be involved in their implementation.*

NAMELESS JUNIOR HIGH SCHOOL

August 15, 19—

Miss Emily Roe
c/o Seaside Resort Hotel
Seaside, New Jersey

Dear Miss Roe:

Starting this September, Nameless Junior High will have a core program in grades 7 and 8, taking the place of English and social studies. Instead of teaching English this year, you will be the seventh-grade core teacher. Our social studies teacher, Mrs. Smith, will teach eighth-grade core.

School begins on Wednesday, September 6. Teachers will report at 9 A.M.

Reprinted from *Bulletin of the National Association of Secondary-School Principals*, February, 1962, 46, 91–103, by permission of the publisher. © Washington, D.C.

Tuesday, September 5, for our first faculty meeting. Tuesday afternoon will be spent getting ready for school opening on Wednesday.

Have a pleasant vacation.

Sincerely,
A. Nonymous, Principal

Thus, like a bolt from the blue, came the word that Miss Roe and Mrs. Smith were, forthwith, to cease teaching specific subject areas and to become "core" teachers. No provisions were made for planning the program ahead of time, developing resource units, gathering teaching materials, or even sitting down with the principal to find out what he meant by "core." Preparation for teaching core was strictly up to Miss Roe and Mrs. Smith in the few weeks before school opened or in whatever time they could spare from daily classroom responsibilities. They did remember hearing core mentioned briefly in some professional education courses, but found their class notes little help in getting ready to teach it. They called upon the librarian for aid in gathering instructional materials, but she was handicapped after school opened because classes were scheduled to use the library nearly every period of the day.

During the first year of the program, Miss Roe's "core" was mostly English, with a little social studies worked in. Mrs. Smith's "core" was the reverse. Students and parents compared notes, and criticized the one-sidedness of the program. By the middle of the second year, however, both teachers were beginning to realize some of the values of an integrated approach. They had located and organized some usable teaching materials, had worked out some procedures that seemed effective, and, in general, felt they were beginning to get their feet on the ground. Imagine their dismay, when it was announced in June that core would be discontinued. The administrator of this school felt that core had failed. In truth, it never had a chance to prove itself.

How often are core programs introduced by administrative fiat, without adequate preparation of all those involved? To what extent are school staff members, students, parents, and the general public involved in orientation to the new program. How much time do schools spend in planning and orientation before initiating a core program? What problems or difficulties are encountered in trying to carry on block-time or core programs? Where core has been tried and dropped, what seem to be the chief reasons for its discontinuance?

Light on these fundamental questions is shed by statewide surveys of block-time, unified studies, and core programs in Alabama,[1] Illinois,[2] Michigan,[3] Minnesota,[4] New Jersey,[5] and Pennsylvania,[6] and a nationwide sampling made by the U.S. Office of Education.[7] Block-time classes are those which meet under one teacher for a block of time embracing two or more regular class periods, and which combine or replace two or more subjects that are required of all pupils and ordinarily taught separately. Subjects are taught separately in most block-time programs, either with or without consciously planned correlation. More properly called unified studies are these block-time classes in which subject matter is unified around a theme, unit, or topic drawn from one of the absorbed subject areas. Core is a block-time class in which content is organized around problems of significance to adolescents, without regard to traditional subject matter boundaries.

---

[1] Otto Hollaway, Scope, trends, and problems of core curriculum work in Alabama, *Educ. Administr. and Superv.*, May 1959, **45**, (3), 153–161.

[2] Illinois Department of Public Instruction, *Block-of-time scheduling practices in Illinois junior high schools.* Springfield, Ill.: The Department, 1960.

Questionnaires sent to all junior, junior-senior, and four-year high schools in the respective states revealed the following percentages of schools having some block-time classes: Alabama 15%,[8] Minnesota 14%,[9] and Pennsylvania 14%,[10]. In Michigan, approximately one half (46%) of the districts operating secondary schools had block-time programs.[11] The New Jersey survey, restricted to junior and junior-senior high schools, identified 51% with block-time,[12] while 61% of Illinois junior high schools had such a program.[13] The nationwide sample revealed block-time in 19.3% of the public junior and junior-senior high schools, double the number found in the 1949 survey.[14]

## Initiating Block-Time, Unified Studies, and Core Programs

Table 1, presenting separate data from several studies, indicates that administrators provided the initiative for a large majority of these programs, often in cooperation with teachers. Only in the Alabama survey do instances

---

[3] Michigan Department of Public Instruction, *The status of block-time programs in Michigan secondary schools,* Bulletin No. 426, Lansing, Mich.: The Department, 1960. 36 pp. (Mimeographed.)

[4] Nelson L. Bossing & John F. Kaufman, Block-time or core practices in Minnesota secondary schools, *The clearing house,* May 1958, 32, (9), 532–537.

[5] Stephen A. Kalapos, *Block-time programs in junior high schools and six-year high schools in New Jersey. Secondary School Bulletin,* No. 2, March 1960, pp. 1–4. Trenton 25: New Jersey State Department of Education.

[6] Pennsylvania Department of Public Instruction. *Report on the 1955–56 Survey of Common Learnings.* Harrisburg, Pa.: The Department, 1958. 16 pp. (Mimeographed.)

[7] U.S. Office of Education, *Block-time classes and the core program in the junior high school,* by Grace S. Wright, Bulletin 1958, No. 6, Washington, D.C.: Government Printing Office, 1958. 70 pp.

[8] Hollaway, *op. cit.,* p. 157.

[9] Bossing, *op. cit.,* p. 532.

[10] Pennsylvania Department of Public Instruction, *op. cit.,* p. 1.

[11] Michigan Department of Public Instruction, *op. cit.,* p. 1.

[12] Kalapos, *op. cit.,* p. 1.

[13] Illinois Department of Public Instruction, *op. cit.,* p. 3.

[14] U.S. Office of Education, *op. cit.,* p. 1.

TABLE 1. How Block-Time Programs Were Initiated[a]

| | Percentage of Schools | | | | |
|---|---|---|---|---|---|
| | Alabama[b] | Illinois[c] | Michigan[d] | Pennsylvania[e] | United States[f] |
| By administrators | 18 | | 37 | | 46 |
| Superintendent | | | | 33 | |
| Principal | | 51 | | 52 | |
| By teachers | 26 | 7 | 7 | 32 | 8 |
| By administrators and teachers | | 33 | 50 | | 33 |
| By the public | | | | 1 | |
| By administrators, teachers, and the public | | | 4 | | 8 |
| Do not know | | | 3 | | |
| No reply to this question | 56 | 9 | | | 5 |
| Total | 100 | 100 | 101[g] | 118[h] | 100 |

[a] Statistics recalculated in some cases to facilitate comparison.
[b] 39 schools in sample. Hollaway, *op. cit.*, p. 158.
[c] 57 schools in sample. Illinois Department of Public Instruction, *op. cit.*, p. 10.
[d] 163 schools in sample. Michigan Department of Public Instruction, *op. cit.*, p. 4.
[e] 143 schools in sample. Pennsylvania Department of Public Instruction, *op. cit.*, p. 3.
[f] Sample composed of 159 schools having unified studies or core. U.S. Office of Education, *op. cit.*, p. 28.
[g] Rounding error.
[h] Some respondents checked more than one item.

of teacher initiative exceed administrators', and here the large number of schools which made no reponse may cast doubt upon the validity of the statistics. The Pennsylvania survey shows principals taking the initiative more often than superintendents. Such data support the contention that the individual school principal is a key factor in curriculum changes of this kind.

## Orientation and Pre-Planning

Some idea of the quality of leadership provided may be inferred from data gathered from several studies shown in Tables 2 and 3. Any important change in curriculum should be preceded by careful pre-planning and orientation of all those who will be affected by the change. Such orientation is especially important in a departure such as core, where content, method, and even the name of the course may be quite different from the usual program. Table 2 indicates that many block-time, unified studies, and core programs were initiated without adequate orientation of all the people involved.

Extensive orientation and planning may not be necessary if a block-time program represents merely a rescheduling of classes so that a teacher has the same pupils for two subjects. The closer a program approximates

TABLE 2. Extent of Orientation and Planning Prior to Starting a Block-Time Program[a]

| | Percentage of Schools | | | |
|---|---|---|---|---|
| | Alabama[b] | Michigan[c] | Pennsylvania[d] | United States[e] |
| Involving staff who would be directly concerned, such as core teachers, guidance counselors, librarian | 21 | 64 | | 47 |
| Entire faculty | 15 | 44 | 51 | 21 |
| A workshop was held | | | 27 | |
| Faculty and students | | | | 3 |
| Students | | 15 | | |
| Students and parents | 8 | | | |
| The public | | 18 | 18 | |
| Faculty and public | | | | 9 |
| Faculty, students and public | | | | 11 |

[a] Percentages do not always total 100% because respondents checked several items, and data is incomplete on some surveys.
[b] Hollaway, *op. cit.*, pp. 158–9.
[c] Michigan Department of Public Instruction, *op. cit.*, p. 5.
[d] Pennsylvania Department of Public Instruction, *op. cit.*, p. 4.
[e] U.S. Office of Education, *op. cit.*, p. 28.

TABLE 3. Amount of Time Provided for Planning and Orientation Prior to Starting the Program

| | Percentage of Schools | | | |
|---|---|---|---|---|
| | Alabama[a] | Michigan[b] | Pennsylvania[c] | United States[d] |
| More than one year | 3 | 12 | 9 | 19 |
| Approximately one year | 13 | 36 | 41 | 43 |
| Three to six weeks | 3 | 15 | | 13 |
| 2 weeks or less | | 17 | | 13 |
| One week | | | 8 | |
| Less than one week | | | 4 | |
| No time | | 2 | | |
| No anwer, or do not know | 82 | 18 | 38 | 13 |
| Total | 101[e] | 100 | 100 | 101[e] |

[a] Hollaway, *op. cit.*, p. 159.
[b] Michigan Department of Public Instruction, *op. cit.*, p. 5.
[c] Pennsylvania Department of Public Instruction, *op. cit.*, p. 4.
[d] U.S. Office of Education, *op. cit.*, p. 29.
[e] Rounding error.

core, however, the greater the amount of time necessary for adequate pre-planning and orientation. Table 3 shows that even when some orientation was provided, too often it amounted to less than two weeks.

### In-Service Help for Core Teachers

Orientation and planning before launching a program are essential, but cannot guarantee a successful block-time program. There must be continuous in-service education and assistance for teachers. The farther the program deviates from the conventional, the greater the need for help and leadership. Difficulties caused by normal teacher turnover are aggravated by the fact that few teachers are prepared specifically for work in block-time programs. Table 4 lists data from several studies showing some of the assistance provided to block-time and core teachers by their administrators. Perhaps it is unrealistic to expect administrators to provide all of the kinds of help listed, but it is apparent that many teachers receive little or no assistance. In fact, on a questionnaire returned by 75 of the 145 Pennsylvania schools having

TABLE 4. Administrative Help Provided for Block-Time Teachers[a]

| | Percentage of Schools | | | |
|---|---|---|---|---|
| | Alabama[b] | Michigan[c] | Pennsylvania[d] | United States[e] |
| In-service training programs—workshops, study groups | 15 | 48 | 35 | 79 |
| Pre-school conferences | | 37 | | |
| Time during school day for cooperative planning by block-time teachers | 3 | 29 | 36 | 46 |
| Time and opportunity for specialists (librarian, art, music, mathematics teachers, *etc.*) to work with block-time classes | 10 | | | 43 |
| Consultants | | 34 | | |
| News sheet or other device whereby block-time teachers share ideas | 8 | | 34 | 31 |
| Special allotment of funds for materials for block-time classes | 5 | 22 | | 36 |
| Source material on the problem | | | 50 | |
| Field trips by block-time classes to utilize community resources | 10 | | | 73 |
| Secretarial help for typing, mimeographing, *etc.* | | | 43 | |
| No specific training or help | | 2 | | |
| No answer | 80 | 13 | | 10 |

[a] Totals exceed 100% because respondents made more than one choice.
[b] Hollaway, *op. cit.*, p. 159–160.
[c] Michigan Department of Public Instruction, *op. cit.*, p. 6.
[d] Pennsylvania Department of Public Instruction, *op. cit.*, pp. 5–6.
[e] U.S. Office of Education, *op. cit.*, p. 43.

block-time in 1957–58, 33 schools (44%) reported they had *no* regular in-service program for these teachers.[15]

The 1955–56 Pennsylvania survey went a step further to sample the kinds of moral support provided by principals. The principals themselves completed most of the questionnaires, so one might expect them to put their best foot forward. On the other hand, allowance must also be made for the possibility that in some schools direct support of the program may have been delegated to an assistant principal, supervisor, or coordinator. Even with these considerations in mind, Table 5 is significant in the number of respondents who, by implication, indicated they did *not* support the program in the ways indicated.

TABLE 5. Moral Support Provided to Block-Time Teachers by the Principal[a]

| | Percentage of Schools[b] |
|---|---|
| Encourages experimentation of teachers | 64 |
| Encourages cooperation with other staff members | 59 |
| Acquaints the teacher with information about available aids | 59 |
| Urges purchase of pertinent materials | 55 |
| Considers the teachers a "guide" rather than a "fountain of knowledge" or a "lesson hearer" | 55 |
| Encourages pupil participation in learning experiences outside the school—community projects, fields trips, *etc.* | 55 |
| Champions the correlation of the subject matter in the program with that of other teachers | 43 |
| Thoroughly orients entire faculty | 38 |
| Encourages visits to other schools using the program | 37 |

[a] Pennsylvania Department of Public Instruction, *op. cit.*, pp. 5–6.
[b] Total exceeds 100% because respondents made more than one choice.

Michigan block-time teachers also indicated the importance of administrative encouragement. When asked what kinds of assistance or resources would help them do a better job, 10% cited "administrative support in terms of encouragement and better awareness of the needs of block-time teachers." Reporting the factors in their present situations which were most helpful in teaching block-time classes, 29% mentioned the administrative climate. Most often cited was freedom to try out new ideas. Terms such as "cooperative," "helpful," "supportive," "sympathetic," and "democratic" were used to describe administrators' commendable personal qualities.[16]

---

[15] Information supplied by Dr. Richard A. Gibboney, Director, Bureau of Curriculum Development, Pennsylvania Department of Public Instruction, Education Building, Harrisburg, Pennsylvania.

[16] Data supplied by Morrel J. Clute, Wayne State University, Detroit, Michigan.

## Obstacles to Block-Time Program Development

Table 6 lists the major obstacles to block-time program development reported by school principals. As might be expected, difficulty in obtaining adequately prepared teachers heads the list. Administrators frequently complain that colleges and universities fail to offer programs and courses for prospective block-time teachers. Yet, some colleges that have offered such programs have been forced to discontinue them for lack of enrollment. Prospective teachers are reluctant to prepare themselves to teach a kind of

TABLE 6. Obstacles to Block-Time Program Development Reported by Principals

| Obstacles | Rank Order | | | | | |
|---|---|---|---|---|---|---|
| | Ala.[a] | Ill.[b] | Mich.[c] | N.J.[d] | Pa.[e] | U.S.[f] |
| Professional Leadership | | | | | | |
| Obtaining adequately prepared teachers | 1 | 1 | 1 | 1 | 1 | 1 |
| Necessity for constant in-service training of teachers | 4 | 7 | 6 | 3 | | 3 |
| Teacher reaction against change | 6 | 3 | 5 | 5 | 2 | 5 |
| Gaining cooperation of other members of the faculty | 6 | 6 | 8 | 6 | | 7 |
| Inability of principal to give sufficient guidance and leadership to the program | | | | | 8 | |
| Public Relations | | | | | | |
| Failure to gain support | | 8 | 9 | 8 | 10 | 9 |
| Certification difficulties | | | | | 7 | |
| Administrative Mechanics | | | | | | |
| Scheduling problems | 6 | 2 | 2 | 2 | 6 | 6 |
| Classes too large | | 5 | | | 3 | |
| Insufficient teacher time for planning | 2 | | 4 | 4 | 4 | 2 |
| Lack of adequate instructional materials | 3 | 4 | 3 | 7 | | 4 |
| Lack of suitable furniture and equipment | | 9 | 7 | | | 8 |
| Lack of proper facilities and instructional materials | | | | | 5 | |
| Lack of help given to teachers, such as assistance in mimeographing and typing, and assistance from the librarian, the audio-visual aids person, and others | | | | | 9 | |

[a] Hollaway, *op. cit.*, p. 160.

[b] Illinois Department of Public Instruction, *op. cit.*, p. 20.

[c] Michigan Department of Public Instruction, *op. cit.*, p. 9.

[d] 66 Schools in sample. Kalapos, *op. cit.*, p. 2.

[e] Pennsylvania Dept. of Public Instruction, *op. cit.*, pp. 6–7. Responses which elaborated on the inadequacies of teachers (Item 1) have been deleted.

[f] U.S. Office of Education, *op. cit.*, p. 48.

course which is not yet widely adopted. Thus we have a vicious circle in which administrators are reluctant to initiate block-time programs for lack of teachers, and teachers are reluctant to prepare for this kind of work for lack of assurance that they can get jobs.

The vicious circle will be broken only as more and more administrators provide the creative leadership necessary to establish successful block-time programs with the teachers available. Only when a sizeable proportion of the secondary schools have some kind of block-time program will large numbers of college students be interested in preparing for this kind of teaching.

Other obstacles cited by principals also point to the key role of the administrator. Although the problem may be more acute with one of the newer programs, it is a prime responsibility of all educational leaders to provide continuous in-service training to increase teacher competence. Such items as dealing with teacher reaction against change and gaining cooperation of other faculty members are familiar aspects of the administrator's role as a change agent. Failure to gain public support is a problem mainly when educational leaders fail to keep the public informed or do not involve broad segments of the community in planning and evaluating the educational program.

Even certification is not completely out of the hands of the local school administrator. Too frequently principals and superintendents assume a fatalistic attitude toward regulations emanating from a state education department. They ought to exert more influence, both individually and through professional associations, to bring about the regulations they think best promote the educational welfare of students.

A number of obstacles to the development of block-time programs are clearly matters of administrative mechanics. One can have nothing but sympathy for the administrator who must make a finite budget and an all-too-limited supply of manpower and facilities cover the almost infinite demands placed upon the schools today. Yet it is to accomplish just such a task that administrative positions have been created. Most administrators are especially selected and trained to wrestle with schedules, and to provide teachers with the equipment, materials, and special services they need. Core programs suffer when such administrative matters are poorly handled.

Where classes are too large, schedules too crowded, and facilities limited, it is the responsibility of the administrative leadership to take the case to the ultimate source of authority in the public schools—the citizens of the community. When the general public sees the value of what is asked, needed funds and facilities nearly always are forthcoming.

## Why Block-Time Programs Were Dropped

Further evidence on the key role of the administrator comes from analysis of the reasons why block-time and core programs were abandoned. The Minnesota survey identified 29 schools that had had a program but abandoned it. Seventeen respondents cited inabiilty to secure qualified teachers; 7 listed scheduling difficulties; and others listed such factors as changes in administration, crowded schools, teacher objections, and lowered scholastic achievement. Only one respondent mentioned community dissatisfaction as a cause for abandoning the program.[17] Lack of qualified personnel also was the major cause of abandonment in Illinois.[18]

The Alabama survey located 25 schools which had dropped their program. Reasons given, in order of frequency, were:

1. Inability to get teachers who are willing and able to carry out this program.
2. Lack of interest due to teachers not being trained for core work.
3. It was scheduled for several years and teachers did not know what to do with it.
4. Teachers continued to teach subjects separately rather than problems or with no conscious correlation.
5. Teachers decided that they could do a more effective job with regular classes.
6. Difficulty of making a schedule and providing time for planning.
7. Teachers are subject matter specialists.[19]

The nationwide survey found that 130 schools from the sample of 2,517 had abandoned their block-time programs. Forty per cent of these gave teacher inadequacies as a reason; 15% listed problems of scheduling; 15% cited opposition from parents, pupils, or teachers; 11% stated that they had failed to get the desired results from the program; 11% cited limitations of space and equipment; 2% mentioned insufficient emphasis on subjects; and 5% gave no reason.[20] Like the obstacles listed in several studies reported in Table 6 above, these difficulties have many implications for administrative leadership, especially the need for effective in-service education.

## Support for Block-Time and Core

Despite the very real difficulties of developing sound block-time and core programs under present conditions, many educators continue to support the idea. For example, the Michigan survey elicited the following responses to the question, "What is your feeling about block-time classes?"[21]

---

[17] Bossing, *op. cit.*, p. 533.

[18] Illinois Department of Public Instruction, *op. cit.*, p. 30.

[19] Hollaway, *op. cit.*, p. 157.

[20] U.S. Office of Education, *op. cit.*, pp. 7–8. Total does not equal 100% due to rounding error.

| | Principals | Teachers |
|---|---|---|
| I strongly favor them | 60% | 52% |
| I feel they have some value | 32% | 35% |
| I have some doubts as to their value | 4% | 9% |
| I do not favor them | 1% | 4% |
| Other, or no response | 3% | 0 |
| Total | 100% | 100% |

It would be difficult to find a better statement of the advantages of block-time classes than the principals' comments summarized in the New Jersey report.

### Advantages of Block-Time Classes Reported by New Jersey Principals

*Teachers have fewer pupils for a longer period of time*

Teachers can get to know individual pupils better—their abilities, needs, interests, strengths and weaknesses, talents.

The guidance function is more readily achieved.

Teachers have a greater opportunity to individualize education to provide for the individual differences among their pupils.

There are greater opportunities for consistent attention to academic, personal, and social development and the development of habits, attitudes, and values.

There are better opportunities for the early identification of pupils with exceptional aptitudes and abilities and the development of individual talent.

There is more constant and sympathetic attention given to continuous growth of the total individual pupil.

*Pupils meet fewer teachers*

A gradual transition from the self-contained elementary school to the departmentalized high school is possible.

There is more effective orientation and adjustment of the pupil to the new junior high-school environment.

Greater security for the pupil accrues by having one or a few teachers know him well.

Better pupil-teacher relations result.

*The same teacher teaches two or more subject areas to one group of pupils*

There is more effective correlation and integration of subject matter areas by teachers and pupils.

---

[21] Michigan Department of Public Instruction, *op. cit.*, pp. 9 and 17.

Greater opportunities exist for application of knowledge and skills—such as language arts skills in social studies learning.

More efficient use of pupil and teacher time is possible by eliminating instructional duplication and concentrating on educational needs.

*Pupils and teacher are together for continuous period of time greater than a single period*

More uninterrupted time allows for greater flexibility and variety in learning activities.

Greater continuity in learning experiences is possible since units or topics may be explored with less regard to fixed time allotments and period changes.

Better opportunities are available for pupil-teacher planning and evaluation.

There is easier scheduling of field trips and the use of resources outside the classroom.

Flexibility in program permits teachers to choose educational topics of value to early adolescents whether or not they "fit" the particular "subject area."[22]

## Conclusion

The evidence here presented places responsibility for a successful block-time, unified studies, or core program squarely upon the shoulders of the administrator. Well-meaning but misguided administrators have given core a bad name in some communities by their ineptness in initiating and guiding the program. On the other hand, there have been instances where successful and well-established programs have been sabotaged or even eliminated overnight by hostile administrators.

Any program which departs materially from the conventional is doomed to failure if it is initiated without adequate orientation and planning, or if those involved are not given continuous support, both material and psychological. Perhaps it is fortunate that inadequate programs are discontinued. Students are apt to be better off in a well-taught conventional program than in a core class taught by a teacher unsympathetic to the idea and ignorant of how to teach it. However, more than 20 years of experience has demonstrated the potential advantages of block-time, unified studies, and core programs. When carefully developed and supported such a program can be a superior vehicle for promoting integrated learning, providing pupil guidance, individualizing instruction, and developing problem-solving and social skills, without sacrificing mastery of essential subject matter.

The many benefits of block-time and core are realized only when administrators practice sound principles of democratic leadership in initiating, developing, and sustaining the program. Pre-service and in-service courses for school administrators should give added stress to these essential skills. Greater attention to these qualifications should be given when administrators are selected. Numerous successful block-time and core programs attest to the fine leadership being provided by many administrators. However, survey results indicate that there is much room for improvement.

---

[22] Kalapos, *op. cit.*, pp. 2–3.

# Part VIII

## *Measurement and Evaluation*

*TWO OF THE most important responsibilities of teachers are assessing the abilities and characteristics of learners and evaluating the extent to which educational objectives have been attained. Educational measurement is the process of obtaining a quantitative representation of the degree to which a trait or ability is possessed by a student. In evaluation, a more inclusive term than measurement, a teacher uses information derived from many sources in order to arrive at a value judgment. The final part in this volume contains six selections pertaining to educational measurement and evaluation.*

*Reading number 44, "Measurement and the Teacher," was authored by Robert Ebel when he was Vice-President for General Programs, at the Educational Testing Service. Ebel discusses ten principles of measurement "selected on the basis of their relevance to the questions and problems which arise most often when tests of educational achievement are being considered, prepared, and used." Brief discussions in support of each principle are also presented.*

*Henry Dyer, Vice-President for College Board Programs, at the Educational Testing Service, discusses the assessment of academic achievement in the following selection. Dyer holds that an educational test is a process that begins and ends in human judgment. Recommended procedures for teacher assessment of academic achievement are based on conceiving of academic achievement in terms of what the student does at the beginning of a course and what he does at the end of the course.*

*Recently, much controversy has surrounded the use and misuse of standardized tests. In reading number 46, Frank Womer selects ten points under which he discusses the misconceptions, misuses, and overuses of standardized test results. Womer is assistant professor of education and consultant on testing and guidance in the Bureau of School Services, University of Michigan.*

*Marvin Glock and Jason Millman, professors of educational psychology and measurement at Cornell University, next discuss the assignment of school marks. The basic point made by the authors is that grading is essentially a value judgment; it is based on numerous measures made by the teacher. The procedures involved in converting these measurements into grades is the substance of their discussion.*

*Reading number 48 is the report of a longitudinal study on the effects of promotion policy on academic achievement. Group treatment of longitudinal data is achieved by using "allometric" charts. The two authors, G. T. Kowitz and C. M. Armstrong, who were associated with the New York State Department of Education at the time their study was published, report some sur-*

prising effects of an "achieve-or-fail" policy. These effects have implications for the mental health of the child.

In the final reading in this part, Irwin Sarason, professor at the University of Washington, reports a study indicating the importance of specific personality variables in academic achievement. Sarason's study differentiates a general anxiety factor from a specific one (i.e., test anxiety) and investigates the relationship between both of them and academic achievement.

# 44

# Measurement and the Teacher — *Robert L. Ebel*

*The author presents ten principles for measuring educational achievement that represent "a sample of the important things classroom teachers need to know about educational measurement." Brief discussions following the statement of each principle help to both clarify and demonstrate the principle's application in school settings. The author feels that "a teacher whose classroom testing reflects an understanding of these principles will do a better than average job of measuring student achievement."*

THE PRINCIPLES of measurement of educational achievement presented in this article are based on the experience and research of a great many people who have been working to improve classroom testing. The particular principles discussed here were selected on the basis of their relevance to the questions and problems which arise most often when tests of educational achievement are being considered, prepared and used. While some of the principles may seem open to question, we believe a case can be made in support of each one.

1. *The measurement of educational achievement is essential to effective education.* Learning is a natural, inevitable result of human living. Some learning would occur even if no special provision were made for it in schools, or no special effort were taken to facilitate it. Yet efficient learning of complex achievements, such as reading, understanding of science, or literary appreciation, requires special motivation, guidance and assistance. Efforts must be directed toward the attainment of specific goals. Students,

Reprinted from *Educational Leadership,* October, 1962, **20,** 20–24, by permission of the author and the Association for Supervision and Curriculum Development.

teachers and others involved in the process of education must know to what degree the goals have been achieved. The measurement of educational achievement can contribute to these activities.

It is occasionally suggested that schools could get along without tests, or indeed that they might even do a better job if testing were prohibited. It is seldom if ever suggested, though, that education can be carried on effectively by teachers and students who have no particular goals in view, or who do not care what or how much is being learned. If tests are outlawed, some other means of assessing educational achievement would have to be used in their place.

2. *An educational test is no more or less than a device for facilitating, extending and refining a teacher's observations of student achievement.* In spite of the Biblical injunction, most of us find ourselves quite often passing judgments on our fellow men. Is candidate A more deserving of our vote than candidate B? Is C a better physician than D? Is employee E entitled to a raise or a promotion on his merits? Should Student F be given a failing mark? Should student L be selected in preference to student M for the leading role in the class play?

Those charged with making such judgments often feel they must do so on the basis of quite inadequate evidence. The characteristics on which the decision should be based may not have been clearly defined. The performances of the various candidates may not have been observed extensively, or under comparable conditions. Instead of recorded data, the judge may have to trust his fallible memory, supplemented with hearsay evidence.

Somewhat similar problems are faced by teachers, as they attempt to assess the achievements of their students. In an effort to solve these problems, tests have been developed. Oral examinations and objective examinations are means for making it easier for the teacher to observe a more extensive sample of student behavior under more carefully controlled conditions.

The price that must be paid for a test's advantages of efficiency and control in the observation of student achievements is some loss in the naturalness of the behavior involved. In tests which attempt to measure the student's typical behavior, especially those aspects of behavior which depend heavily on his interests, attitudes, values or emotional reactions, the artificiality of the test situation may seriously distort the measurements obtained. But this problem is much less serious in tests intended to measure how much the student knows, and what he can do with his knowledge. What is gained in efficiency and precision of measurement usually far outweighs what may be lost due to artificiality of the situation in which the student's behavior is observed.

3. *Every important outcome of education can be measured.* In order for

an outcome of education to be important, it must make a difference. The behavior of a person who has more of a particular outcome must be observably different from that of a person who has less. Perhaps one can imagine some result of education which is so deeply personal that it does not ever affect in any way what he says or does, or how he spends his time. But it is difficult to find any grounds for arguing that such a well concealed achievement is important.

If the achievement does make a difference in what a person can do or does do, then it is measurable. For the most elementary type of measurement requires nothing more than the possibility of making a verifiable observation that person or object X has more of some defined characteristic than person or object Y.

To say that any important educational outcome is measurable is not to say that satisfactory methods of measurement now exist. Certainly it is not to say that every important educational outcome can be measured by means of a paper and pencil test. But it is to reject the claim that some important educational outcomes are too complex or too intangible to be measured. Importance and measurability are logically inseparable.

4. *The most important educational achievement is command of useful knowledge.* If the importance of an educational outcome may be judged on the basis of what teachers and students spend most of their time doing, it is obvious that acquisition of a command of useful knowledge is a highly important outcome. Or if one asks how the other objectives are to be attained —objectives of self-realization, of human relationship, of economic efficiency, of civic responsibility—it is obvious again that command of useful knowledge is the principal means.

How effectively a person can think about a problem depends largely on how effectively he can command the knowledge that is relevant to the problem. Command of knowledge does not guarantee success, or happiness, or righteousness, but it is difficult to think of anything else a school can attempt to develop which is half as likely to lead to these objectives.

If we give students command of knowledge, if we develop their ability to think, we make them intellectually free and independent. This does not assure us that they will work hard to maintain the status quo, that they will adopt all of our beliefs and accept all of our values. Yet it can make them free men and women in the area in which freedom is most important. We should be wary of an educational program which seeks to change or control student behavior on any other basis than rational self-determination, the basis that command of knowledge provides.

5. *Written tests are well suited to measure the student's command of useful knowledge.* All knowledge can be expressed in propositions. Propositions are statements that can be judged to be true or false. Scholars, scientists, re-

search workers—all those concerned with adding to our store of knowledge, spend most of their time formulating and verifying propositions.

Implicit in every true-false or multiple-choice test item is a proposition, or several propositions. Essay tests also require a student to demonstrate his command of knowledge.

Some elements of novelty are essential in any question intended to test a student's command of knowledge. He should not be allowed to respond successfully on the basis of rote learning or verbal association. He should not be asked a stereotyped question to which a pat answer probably has been committed to memory.

6. *The classroom teacher should prepare most of the tests used to measure educational achievement in the classroom.* Many published tests are available for classroom use in measuring educational aptitude or achievement in broad areas of knowledge. But there are very few which are specifically appropriate for measuring the achievement of the objectives of a particular unit of work or of a particular period of instruction. Publishers of textbooks sometimes supply booklets of test questions to accompany their texts. These can be useful, although all too often the test questions supplied are of inferior quality—hastily written, unreviewed, untested, and subject to correct response on the basis of rote learning as well as on the basis of understanding.

Even if good ready-made tests were generally available, a case could still be made for teacher-prepared tests; the chief reason being that the process of test development can help the teacher define his objectives. This process can result in tests that are more highly relevant than any external tests are likely to be. It can make the process of measuring educational achievement an integral part of the whole process of instruction, as it should be.

7. *To measure achievement effectively the classroom teacher must be (a) a master of the knowledge or skill to be tested, and (b) a master of the practical arts of testing.* No courses in educational measurement, no books or articles on the improvement of classroom tests, are likely to enable a poor teacher to make good tests. A teacher's command of the knowledge he is trying to teach, his understanding of common misconceptions regarding this content, his ability to invent novel questions and problems, and his ability to express these clearly and concisely; all these are crucial to his success in test construction. It is unfortunately true that some people who have certificates to teach lack one or more of these prerequisites to good teaching and good testing.

However, there are also some tricks of the trade of test construction. A course in educational measurement, or a book or article on classroom testing can teach these things. Such a course may also serve to shake a teacher's faith—constructively and wholesomely—in some of the popular misconceptions about the processes of testing educational achievement. Among these misconceptions are the belief that only essay tests are useful for measuring the development of a student's higher mental processes; that a test score should indicate what proportion a student does know of what he ought to know; that mistakes in scoring are the main source of error in test scores.

8. *The quality of a classroom test depends on the relevance of the tasks included in it, on the representativeness of its sampling of all aspects of instruction, and on the reliability of the scores it yields.* If a test question presents a problem like those the student may expect to encounter in his later life outside the classroom, and if the course in which his achievement is being tested did in fact try to teach him how to deal with such problems, then the question is relevant. If the test questions involve, in proportion to their importance, all aspects of achievement the course undertakes to develop, it samples representatively. If the scores students receive on a test agree closely with those they would receive on an independent, equivalent test, then the test yields reliable scores.

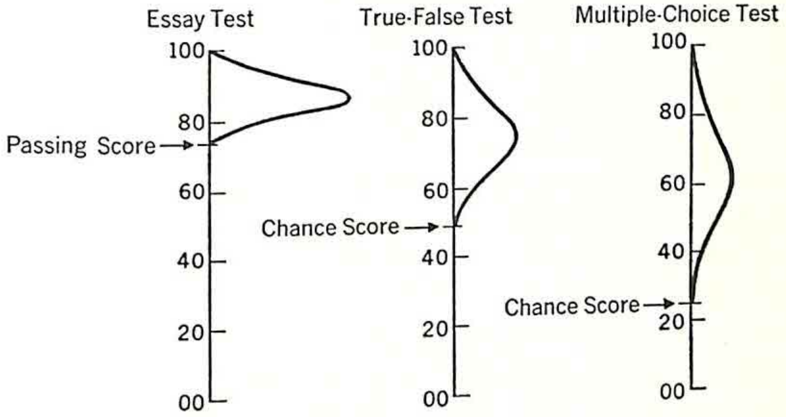
Relevance, representativeness and reliability are all matters of degree. Procedures and formulas for calculating estimates of test reliability are well developed, and are described in most books on educational measurement. Estimates of representativeness and relevance are more subjective, less quantitative. Yet this does not mean that relevance and representativeness are any less important than reliability. The more a test has of each the better. While it is possible to have an irrelevant and unrepresentative but highly reliable test, it is seldom necessary and never desirable, to sacrifice any one of the three for the others.

Either essay or objective test forms can be used to present relevant tasks to the examinees. Ordinarily, the greater the novelty of a test question, that is, the smaller the probability that the student has encountered the same question before, or been taught a pat answer to it, the greater its relevance. Because of the greater number of questions involved, it is sometimes easier to include a representative sample of tasks in an objective than in an essay test. For the same reason, and also because of greater uniformity in scoring, objective tests are likely to yield somewhat more reliable scores than are essay tests.

9. *The more variable the scores from a test designed to have a certain maximum possible score, the higher the expected reliability of those scores.* Reliability is sometimes defined as the proportion of the total variability among the test scores which is not attributable to errors of measurement. The size of the errors of measurement depends on the nature of the test—the kind and the number of items in it. Hence for a particular test, any increase in the total variability of the scores is likely to increase the propor-

tion which is not due to errors of measurement, and hence to increase the reliability of the test.

Figure 1 shows some hypothetical score distributions for three tests. The essay test consists of 10 questions worth 10 points each, scored by a teacher who regards 75 as a passing score on such a test. The true-false test consists of 100 items, each of which is worth one point if correctly answered, with no subtraction for wrong answers. The multiple-choice test also includes 100 items, each of which offers four alternative answer options. It, too, is scored only for the number of correct answers given, with no "correction for guessing."



| | Essay Test | True-False Test | Multiple-Choice Test |
|---|---|---|---|
| Mean | 87.5 | 75 | 62.5 |
| Standard Deviation | 5 | 10 | 15 |
| Effective Range | 25 | 50 | 75 |
| Estimated Reliability | .50 | .67 | .80 |

FIG. 1. *Hypothetical Score Distributions for Three Tests.*

Note, in the data at the bottom of Figure 1, the differences among the tests in average score (mean), in variability (standard deviation), in effective range and in estimated reliability. While these are hypothetical data, derived from calculations based on certain assumptions, they are probably reasonably representative of the results most teachers achieve in using tests of these types.

It is possible to obtain scores whose reliability is above .90 using 100 multiple-choice items, but it is not easy to do, and classroom teachers seldom do it in the tests they construct. It is also possible to handle 100-point essay tests and 100-item true-false tests so that their reliability will equal that of a 100-item multiple-choice test. But again, it is not easy to do and classroom teachers seldom succeed in doing it.

10. *The reliability of a test can be increased by increasing the number of questions (or independent points to be scored) and by sharpening the power of individual questions to discriminate between students of high and low achievement.* Figure 2 illustrates the increases of test reliability which can be expected as a result of increasing the number of items (or independent points to be scored) in a test. Doubling the length of a 10-item test whose reliability coefficient is .33 increases the reliability to .50. Doubling again brings it up to .67, and so on. These estimates are based on the Spearman-Brown formula for predicting the reliability of a lengthened test. While the formula requires assumptions which may not be justified in all cases, its predictions are usually quite accurate.

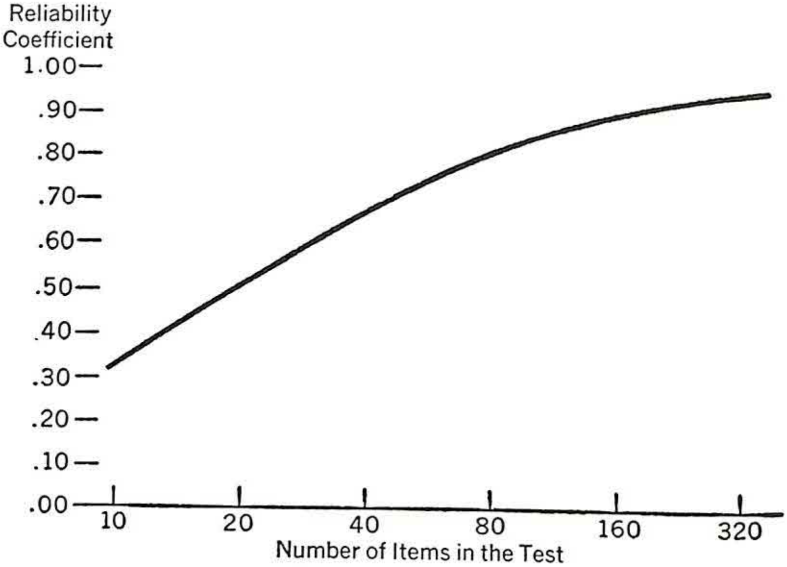


FIG. 2. *Relation of Test Reliability to Test Length.*

Figure 3 shows how the maximum discriminating power of an item is related to its level of difficulty. These discrimination indices are simply differences between the proportions of correct response from good and poor students. Good students are those whose total test scores fall among the top 27 percent of the students tested. Poor students are those whose scores make up the bottom 27 percent. An item of 50 percent difficulty does not necessarily have (and usually will not have) an index of discrimination of 1.00. Its discriminating power may be zero, or even negative. But items of middle difficulty have higher ceilings on their discriminating power. What is more important, they not only can have, but usually do have, greater discriminating power than very easy or very difficult items. An item that no one answers correctly, or that everyone answers correctly, cannot discriminate at all. Such an item adds nothing to the reliability of a test.

In summary, the 10 principles stated and discussed in this article represent only a sample of the important things classroom teachers need to know about educational measurement. These principles, and the brief discussion of each presented here, may serve to call into question some common practices in classroom testing, or to suggest some ways in which classroom tests might be improved. They are not likely, and are not intended, to say all that needs to be said or do all that needs to be done to improve educational measurement in the classroom. It is our sincere belief, however, that a teacher whose classroom testing reflects an understanding of these principles will do a better than average job of measuring student achievement.

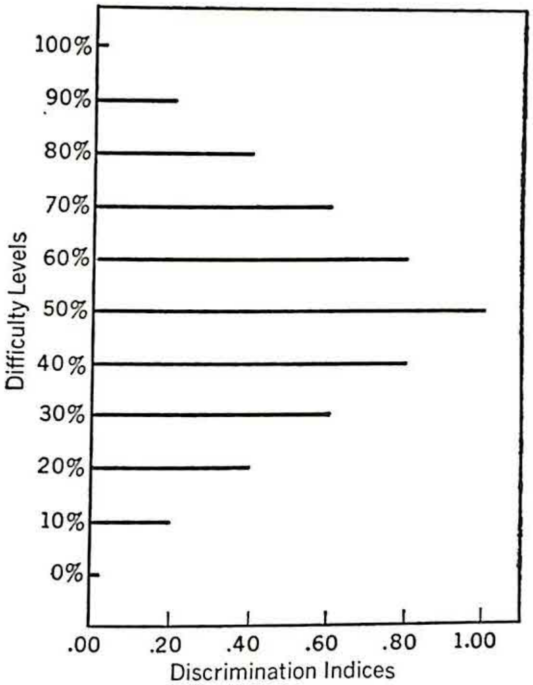


FIG. 3. *Maximum Discrimination Attainable with Items at Different Levels of Difficulty.*

# 45

# On the Assessment of Academic Achievement

## — *Henry S. Dyer*

*The assessment of student achievement is viewed in the present article as having central importance in the educational enterprise. Classroom teachers, administrators, and researchers at all levels rely on numerous testing devices to assess academic achievement. The author discusses some of the essential points in the testing process. Emphasizing that it is primarily a human process, the author goes on to define academic achievement and the teacher's role in its assessment.*

IN THESE days of ferment in the curriculum, any large producer of educational tests is likely to get strange and wonderful requests from teachers suddenly smitten by the need to evaluate the new things they are doing. The following letter, though purely imaginary, is not untypical. It reflects a few of the mistaken ideas about testing and the assessment of achievement which are prevalent:

Dear Sir:

During this past academic year I have been giving a new course, Marine Wild Life 106. I need some incontrovertible evidence to prove to my suspicious colleagues that Marine Wild Life 106 is more comprehensive and effective than the course it has replaced, namely, Marine Wild Life 102.

Will you therefore please send me 92 copies of your standardized objective test in ichthyology? I am particularly interested in showing that my students

Reprinted from *Teachers College Record*, 1960, 62, 164–172, by permission of Henry S. Dyer, Vice President, Educational Testing Service, and the publisher.

have developed a true and deep appreciation of the love-life of our underwater friends.

Please rush the shipment, since I want to give the test as part of my final examination next Thursday.

Sincerely yours,
Professor Finn

Although Professor Finn does not exist, he has many educational cousins who do, and who think, with him, that the purpose of educational evaluation is to demonstrate the truth of a foregone conclusion. It has never occurred to him that an investigation of Marine Wild Life 106 should be so planned as to permit negative findings to appear if, in fact, the course is not all he firmly believes it to be.

But what most bothers the professional tester about the many Professor Finns who are experimenting with new courses is the series of bland assumptions that lie behind the letters they write. The first such assumption is that one can prove something by giving a single test after the show is all over. The second is that a standardized objective test is standard and objective in some absolute sense, as though it had been made in heaven. The third is that testing agencies are like vending machines: All you have to do is put your nickel in the right slot, and out will come precisely the test you are looking for. All of which would be just dandy if true, but unfortunately none of these assumptions bears any noticeable relation to the facts of testing and the assessment of achievement.

## The Human Side of Tests

What Professor Finn does not appear to realize is that an educational test of any kind is primarily a human process, not a physical thing. It is a process that begins and ends in human judgment. It is a process that requires time, effort, and hard creative thinking, not only of the professional tester, but of the test user as well. It is a process that can take an infinite number of forms, depending on the purposes to be served and the conditions under which the testing is to be done. It is a process that can hardly be confined to the use of multiple choice questions and paper-and-pencil techniques if it is to provide the sort of rewarding illumination of the educational scene that some people hope for.

In what follows I shall discuss some of the essentials of the testing process as applied to the assessment of academic achievement and pay my repect to several of the knotty problems that bedevil the whole enterprise. The discussion will turn up more questions than answers, but if the questions are sufficiently disturbing, the outcome should be such as to nudge Professor Finn toward deeper wisdom in approaching his evaluation problem.

What do we mean by "academic achievement?" If it is something we are trying to assess, then it seems reasonable to ask first of all what it is. In current usage, it is a fuzzy term that may mean any one of a dozen unspecified things: the sum total of information a student has at his command when he finishes a course of instruction, the getting of a passing grade in a course regardless of what may lie behind the grade, the score on a test that has "achievement" in the title, and so on.

There are two ideas that can be used to pin down the notion of academic achievement a bit more precisely. The first idea is that academic achievement refers to the *identifiable operations* a student is expected to perform on the materials of a course, that is, on the facts, theories, problems, principles, and points of view which he encounters while taking the course. The second idea is that academic achievement refers to the *differences* between the number and kinds of operations a student can and does perform at the beginning of a course and the number and kinds of operations he can and does perform at the end of a course.

The key terms in this definition of achievement are *identifiable operations* and *differences*. The emphasis on operations is supposed to suggest that it is what the student actually *does* that counts. We should think of achievement as something composed of transitive verbs with direct objects—verbs like *infer, generalize, recall, compare, analyze, evaluate, organize, criticize.* And we need to be sure that we can identify these operations by reference to specific tasks or questions that require them. ("What inferences can you draw from the following set of data?" "What general principle can you find that explains the behavior of the following political figures?" etc.)

The emphasis on the *differences* between what a student does at the beginning of a course and what he does at the end of a course calls attention to the fact that academic achievement is a dynamic, not a static, concept; it is what has happened between then and now, not just what is happening now. Can the student now solve differential equations that he found impossible last October? Can he detect differences between Beethoven and Brahms that were not apparent to him earlier? Can he organize his thoughts in written form better than he did at the outset?

Clearly, under this definition, the assessment of academic achievement is a complicated business; it is not something that can be done merely by going out and buying a standardized test two weeks before the final examination. It requires that the teacher responsible for a course must himself be a major participant in the assessment job from start to finish. He may be able to get supplementary aid from published tests; he may find that the professionals in testing can give him useful ideas and point the way to sound strategy; but the substance of the job of sizing up what he has done to his students rests with him.

How shall he proceed?

## Clarity of Goals

Obviously, the first step is for him to get a clear idea of what he is trying to have his students achieve. What operations does he want them to perform in June that they could not perform last September? This may be an "obvious" first step, but a look around the classrooms of the country suggests that it is a step rarely taken. Almost invariably teachers are more concerned with what *they* are going to do in their courses than with what their *students* are going to do. Lessons and lectures are outlined, reading lists are laid out, visual aids are planned, but few teachers give much concentrated attention to what the fuss is really all about from the standpoint of the student. If a syllabus is prepared, it almost never gives any detailed description of what students are supposed to be able to do as a result of having been exposed to the instructor and the subject matter. There is usually a hope that students will be impressed, that they will remember something, and in a vague sort of way it is felt that the course experiences will do them good, that is, make better thinkers or better citizens of them. But there is very little disposition to get down to brass tacks and specify what kinds of student performance are deemed to reflect better thinking or better citizenship.

All of which is hardly surprising. Stating the goals of a course in a highly concrete manner takes more energy and imagination than most teachers have time for. One of the very practical problems in developing a sound assessment program is to find the time needed to carry out this essential first step.

Generally speaking the goals of achivement fall into three broad classes: informational goals, proficiency goals, and attitudinal goals. The informational goals refer to those items of information that students are expected to know and give forth on demand by the time they have completed a course. It would appear that these can be readily described simply by listing the topics, sub-topics, and sub-sub-topics the course is expected to cover. But this is only half the story. A fact in human knowledge is not just the name of something standing all by itself; it is a subject and a predicate. It is not just the name "sodium"; it is the sentence, "Sodium is a metal." Progress toward the informational goals of a course, therefore, is measured by the number and complexity of these subject-predicate relationships which the student can reproduce. In making the informational goals useful as a basis for assessing achievement, some explicit attention must be given to these relationships and to their degree of complexity. It is not sufficient to list topics and sub-topics and sub-sub-topics.

The large majority of teachers would disown the idea that the *only* kinds of goals they have in mind for their courses are informational goals. They would argue, and quite properly, that information is one of the goals, and a necessary one, but certainly not one of ultimate importance. This they argue, but their arguments are too often confounded by their own examinations, which demand facts and facts only. One of the hoariest criticisms brought against objective tests is that they test only for factual information, while essay tests get at the higher aspects of learning. This would be a cogent criticism except for two things: First, a well-made objective test need not be limited to the measurement of factual information; and second, essay tests are too frequently graded solely by counting the number of relevant facts that each student churns up.

Most of what we think of as "the higher aspects of learning" is contained in the proficiency goals of academic achievement, that is, proficiency in various kinds of skills, both manual and mental–manipulative skills, problem solving skills, evaluative skills, and organizational skills. Taken together, they add up to effective thinking and sound execution. But to be useful as guides in the assessment of achievement, they must *not* be taken together; they must be elaborated in detail and expressed in terms of the multitude of different kinds of specific tasks a student is expected to perform and in terms of the specific kinds of operations he must follow in performing them.

Unless the proficiency goals are elaborated in this fashion, they may be too easily overlooked or forgotten. Back in the 1930's a number of experiments in the teaching of science led to the conclusion that science courses without laboratory instruction were just as effective as science courses with laboratory instruction. The typical set-up was to divide science classes into two groups, one of which received laboratory work in addition to the regular course work and the other of which received no laboratory exercise. At the end of the course, the two groups would be compared by means of an achievement test in science. The differences in test scores between the two groups were generally negligible. Why? Because in setting up the tests, the proficiency goals peculiar to laboratory instruction had been overlooked. The only measures used had been measures related mainly to informational goals that have little or nothing to do with what goes on in a laboratory. Under these conditions, the finding of no difference is essentially meaningless, since the relevant proficiency variables had not been taken into account. More recently Kruglak and his associates at Minnesota have taken a second look at this problem in science teaching, have devised tests aimed directly at the proficiency goals of laboratory instruction, and have come up with results of a more meaningful sort (7).

It is admittedly no easy matter to define all the proficiency goals in terms of student performance. One can flounder a long time over the types of performance that form the components of creative thinking, for instance, or skill in evaluative judgment. The result is that nobody has yet been able to produce a completely satisfactory and universally applicable method for assessing achievement in these areas. On the other hand, there have been some interesting attacks on the problem, and possibly in the course of time it may be solved (*8*, *10*, *11*). A giant step toward the solution will be made when teachers now uttering pious sentiments on the subject will get down to cases and try to define what they are really looking for when they speak of such things as "creative power" and "sensitivity to values."

Finally, there are the attitudinal goals—those educational objectives which are often blithely called the "intangibles." Everybody applauds them, but few teachers seem to know what to do about them. How does one define such goals as love of music or sense of social responsibility or enthusiasm for abstract ideas in such a way that they can be recognized when seen? What is it that a student does to demonstrate that he likes good literature? We have our dodges on matters of this sort: We count the books a student takes out of the library; we measure the amount of spare time he gives to good reading (provided, of course, we allow him any spare time at all); we engage him in conversation and try to judge from the fervor in his voice how far along the scale of appreciation his reading has brought him, or we ask him outright on some sort of rating schedule how well he likes Shakespeare and T. S. Eliot. But we suspect that these devices for uncovering attitudes are based on exceedingly tenuous inferences. Books may be taken out of the library not to be read, but to maintain an impressive looking bookshelf. The time spent in reading may really be spent only in an effort to avoid reality. Conversational fervor may be only good acting or the effect of one cocktail too many. And a rating schedule can almost always be faked if much depends upon it.

I dislike to be discouraging in this all-important matter of attitudinal goals, but thus far very few operations have been suggested which define them satisfactorily. There are some breaks in the clouds, though. At Pennsylvania State University a group of experimenters has been trying to get a line on student attitudes toward television instruction. One device they have hit upon is to give the students generous samples of both TV instruction and conventional instruction, and then to permit them a genuine choice as to which type they would take for the remainder of the semester (3). This is a real operational definition of a specific attitude. It furnishes a fruitful clue for further development in defining attitudes in other areas, the general principle being that one defines attitudes in terms of live decisions that make an actual difference in what an individual will do or not do.

## Measurements of Goals

Once goals of instruction have been clearly set forth by describing the operations students have to perform to attain them, the problem of devising tests and other techniques for assessing academic achievement is at least seventy-five per cent solved. The reason is that an achievement test is in effect a sample of all the kinds of tasks that a given course of study is striving to get students to master. As such, the tests should in themselves constitute a definition of the goals to be attained. In putting together an appropriate sample of tasks and checking out their adequacy, there are technical matters in which the experienced tester may serve as a guide to the teacher. But, again, the questions of what to include or not to include, what is more important and what is less important, what constitutes a good response and what constitutes a poor response must in the last analysis be the teacher's decision if the test is to measure progress toward the goals *he* has in mind.

On the other hand, as mentioned above, teachers are busy people who are unlikely to find the time they need to do all that really should be done to work out the goals of their own instruction and to devise fully valid measures for assessing achievement. What are some of the practical compromises?

In the first place, although one is unlikely to find any published standardized test that in all respects fits the goals of a particular course or school, one may, by analyzing such tests question by question, possibly find one that comes reasonably close. Looking at a test question by question is important not only for determining a test's general suitability, but also for deciding what questions might properly be eliminated in assessing the performance of one's own students. Surprisingly, the notion of adapting a published test by dropping out irrelevant material seldom occurs to people, yet in many circumstances it is an obvious and justifiable procedure.

Standardized tests, when they can be found and adapted, have two important advantages for any program for assessing academic achievement. First, the makers of such tests have usually lavished a great deal of care on the preparation and tryout of each question to make sure that it is unambiguous, that it discriminates sharply between good and poor students, and that it is of the right level of difficulty for the group at which it is aimed. Second, two or more parallel forms of a standardized test are usually available, so that one has at his disposal the means for accomplishing a highly important part of the assessment job, that of measuring the student's performance twice—one at the beginning of the course and once again at the end.

### Testing Thinking Ability

Until recently most standardized tests have concentrated rather heavily on the purely informational goals of instruction and have tended to neglect the proficiency goals. In the last few years, however, there has been a marked change in this respect. More and more tests are made up of questions that require the student to perform various mental operations well beyond the simple operation of recalling learned facts. They are getting at problem solving and reasoning processes of many kinds. Consider the following sets of questions drawn from a booklet describing the achievement tests of the College Entrance Examination Board (*1*). If the reader will take the trouble to wrestle with them, I think he will find that they challenge powers of thought well beyond simple memory, though they also require of the student a firm foundation of factual knowledge.

The first set of questions is intended to test the student's appreciation of those factors that must be controlled so that valid conclusions can be drawn from the results of a scientific experiment:

You are to conduct an experiment to determine whether the rate of photosynthesis in aquatic plants is affected by the addition of small amounts of carbon dioxide to the water in which the plants are growing. You have available all of the equipment and material generally found in a well-equipped science laboratory plus several well-rooted Elodea plants which are growing in a battery jar. It is evident that the plants are healthy because they are giving off bubbles of oxygen.

1. If the addition of carbon dioxide were to have an effect upon aquatic plants, almost immediately after bubbling the gas through the water you would expect to observe a noticeable change in the
   (A) growth of the plants
   (B) temperature of the water
   (C) coloring of the plant leaves
   (D) rate of bubble production by the plant
   (E) amount of oxygen consumed by leaf respiration
2. If a carbon dioxide generator were not available for the experiment, an adequate supply of carbon dioxide might be provided by
   (A) placing a piece of carbon in the water in which the plants are growing
   (B) burning a candle over the battery jar
   (C) adding carbon tetrachloride to the water
   (D) pumping air into the water
   (E) blowing through a glass tube into the water
3. Which of the following materials would be *least* useful in identifying the escaping gas as oxygen?

(A) A collection bottle
(B) Glass tubing
(C) A glass plate
(D) Splints
(E) Matches

4. Which of the following materials could be used to provide a source of carbon dioxide for this experiment?
(A) Distilled water and carbon
(B) Zinc and carbonic acid
(C) Limestone and hydrochloric acid
(D) Limewater and carbon
(E) Sodium hydroxide and carbon

5. The addition of carbon dioxide could have no observable effect upon the rate of bubbling by the Elodea plants if you had
(A) previously turned off the light
(B) no means of regulating the carbon dioxide flow
(C) no means of regulating the room temperature
(D) a variable water temperature
(E) no cover on the battery jar

6. Before you could conclude whether or not the rate of photosynthesis in aquatic plants in general is affected by the addition of small amounts of carbon dioxide to the surrounding water, you would have to repeat the experiment using
(A) all the materials of the original experiment
(B) an entirely different set of materials
(C) other aquatic plants
(D) other Elodea plants
(E) some terrestrial plants

The following set of questions attempts to test the student's knowledge and understanding of United States foreign policy:

SPEAKER I: I don't think the United States has any business getting involved in the affairs of other nations, through foreign aid or any thing else. We have prospered without aid from other countries. Why can't other nations do the same?

SPEAKER II: But we can't afford not to aid other nations. A foreign-aid program offers us a great opportunity to increase our prestige and we should take advantage of it. A nation which ignores other nations will not be regarded as important.

SPEAKER III: Maybe, but let's not forget the fact that foreign aid is also an investment in our own future. If we don't help other free nations, we can't expect to stay free ourselves.

SPEAKER IV: Let's be practical about this. If other nations are too weak to stand on their own two feet, we should help them, yes, but let's remember that when they become dependent on us we must also subordinate them to us. We can only justify foreign aid if we're going to protect our investment.

SPEAKER V: You're all so cold-blooded about this! It isn't a matter of practicality, but of moral obligation to aid other countries. If we could only renounce war, think of the money and material which would be available for constructive work.

1. A humanitarian point of view is best represented by Speaker (A) I (B) II (C) III (D) IV (E) V
2. Since the close of World War II, the United States has mainly justified its foreign-aid policy by arguments such as those advanced by Speaker (A) I (B) II (C) III (D) IV (E) V
3. Which speaker represents a point of view historically associated with the midwestern United States?
   (A) I (B) II (C) III (D) IV (E) V
4. Speaker IV would probably have approved of the
   1. Platt Amendment
   2. Roosevelt Corollary to the Monroe Doctrine
   3. Good Neighbor Policy
   (A) 3 only
   (B) 1 and 2 only
   (C) 1 and 3 only
   (D) 2 and 3 only
   (E) 1, 2, and 3

* * *

Despite the progress that may have been made in developing standardized tests, one must always bear in mind that they can never do the whole job of assessing academic achievement in a particular situation. In the last analysis, if a teacher wants to measure those aspects of achievement peculiar to his own instruction he must resort to at least some tests of his own making. He can get some good ideas for approaching this enterprise from two books: *Taxonomy of Educational Objectives* (2) and *General Education: Explorations in Evaluation* (5). If one peruses these books looking for new testing ideas rather than faults, one may find much that is helpful in suggesting ways and means of measuring progress toward the goal of more effective thinking.

But the assessment of the student's growth in effective thinking is still not enough. Any teacher with a well-developed conscience is primarily concerned that his students shall acquire increasing maturity in their attitudes, that is, a more grown-up and informed approach to life and learning. The assessment of this aspect of achievement requires ingenuity and a willingness to experiment with tests that might be regarded as "off-beat."

## Testing Maturity

Two such instruments may be mentioned as examples. One, called *The Facts About Science* (9), looks on the surface like an ordinary factual test aimed at seeing how much the student knows about scientists and the sorts of things they do. Actually, however, the pay-off response to each of the scored questions represents a false stereotype of the scientist or his work, e.g., "A scientist has no sense of humor." The test has been used to see to what extent high school courses in science influence students away from these false stereotypes. The same approach could be taken to get at student attitudes toward many other social institutions and enterprises.

Another instrument carries the title *Sizing Up Your School Subjects* (6). It focuses on students' attitudes toward their current academic work and is especially designed for use in educational experiments where a given subject is being taught by two or more methods. Through an indirect approach, which almost certainly guarantees that the student will not fake his responses, it provides the means for comparing different courses or different methods of instruction in the same course with respect to the reactions of students to ten aspects of the material or the teaching. (E.g., What course is best at holding the student's attention? What course is regarded by the student as most valuable to him?) With a little imagination, the technique can be elaborated to cover the attitudes of students toward almost any feature of the academic work in which they are engaged.

## Alternative Devices

One of the things that is holding back the development of new and better methods of assessing academic achievement, especially in the area of attitudes, is the tendency to think almost exclusively in terms of paper-and-pencil tests, particularly multiple choice tests. Multiple choice tests are useful up to a point. They can do more than most people realize in tapping a student's higher mental processes, and all the possibilities of multiple choice questions have not yet been fully explored. Nevertheless, the multiple choice test has, by its very nature, certain severe limitations. So do other forms of written tests, including essay tests.

To break out of this restricted mode of thinking we ought to consider how to exploit the possibilities in the so-called situational tests which psychologists have been working on in recent years. To what extent can such tests be adapted to the measurement of the "intangibles" of academic achievement? Situational tests are still highly experimental and full of technical problems, but they may well be the answer to the question of how to evaluate many of the subtle aspects of human behavior that can never be reached satisfactorily by paper-and-pencil devices.

One effort to apply a situational test in the classroom setting can be found in *The Russell Sage Social Relations Test* (4). On the surface, this is a test of the ability of a group of students to solve cooperatively a series of construction problems. It does in fact give a reasonably good measure of the group's effectiveness in this situation. If properly administered, however, it also yields information about other important aspects of group behavior: how well the students have learned to work together, how they respond to group control, how well they develop among themselves an efficient group organization, how they react on one another, and so on.

This sort of technique, and others like it, should eventually bring us closer to a truly adequate assessment of how students are attaining the attitudinal goals of instruction. Indeed, such techniques can do more. They can dramatize the importance of the goals by demonstrating how far we are falling short of achieving them. The development and application of good situational tests are expensive and time-consuming. Nevertheless, if we really wish to know what is happening to students in our classrooms, we are going to have to give up the notion that we can find out by relying on the "quick-and-dirty" tests that come cheap.

## Assessment Is Central

Education these days has become a major concern of national policy. It is in consequence the center of a good deal of bitter controversy, and the issues being debated (segregation, the source of school support, teachers' salaries, federal or local control, "standards") are so overriding that the question of how to assess student achievement seems by comparison to be of minor technical importance. Actually, it ought to be regarded as of central importance to the whole educational enterprise. Until we can reduce our vast ignorance of what is actually happening to the minds and hearts of students in the classrooms, until we can point up the goals of education in terms of what we want students to *do* physically, mentally, spiritually, and until we have better ways of knowing how well we are getting them to do these things, there is hardly much point to all the fuss about the so-called "larger issues." A comprehensive program to develop better methods for assessing achievement in all the classrooms of the country would go a long way toward taking the heat out of the controversies now plaguing us and would furnish the means for replacing the anarchy in education with new vitality and a sense of direction.

## REFERENCES

1. *A description of the College Board Achievement Tests.* Princeton, N. J.: College Entrance Examination Board, 1960.
2. Bloom, B. S. (Ed.) *Taxonomy of educational objectives.* New York: Longmans, Green and Co., 1956.
3. Carpenter, C. R., & Greenhill, L. P. *An investigation of closed circuit television for teaching university courses.* University Park, Pa.: The Pennsylvania State University, 1958. P. 77.
4. Damrin, D. E. The Russell Sage Social Relations Test: a technique for measuring group problem solving skills in elementary school children. *J. of exp. educ.,* 1959, **28,** 85–99.
5. Dressel, P. L., & Mayhew, L. B. *General education: Explorations in evaluation.* Washington, D.C.: American Council on Education, 1954.
6. Dyer, H. S., & Ferris, A. H. *Sizing up your school subjects.* Princeton, N. J.: Educational Testing Service, 1958.
7. Kruglak, H., & Carlson, C. R. Performance tests in physics at the University of Minnesota. *Science Education,* 1953, **37**(2), 108–121.
8. MacKinnon, D. W. The highly effective individual. *Teachers Coll. Rec.,* 1960, **61,** 367–378.
9. Stice, G. *The facts about science test.* Princeton, N. J.: Educational Testing Service, 1958.
10. Taylor, C. W. (Ed.) *Research conference on the identification of creative scientific talent.* Salt Lake City, Utah: University of Utah Press, 1956.
11. Wilson, R. C. Improving criteria for complex mental processes. *Proceedings of the 1957 Invitation Conference on Testing Problems.* Princeton, N. J.: Educational Testing Service, 1957.

# 46

# Testing Programs — Misconceptions, Misuse, Overuse — *Frank B. Womer*

*There has been a rapid growth of standardized testing at all grade levels. The present author attributes this to a "natural growth" as well as to the rapid growth of the guidance movement. The ultimate criterion for judging the effectiveness of testing programs is the correct use of test results. In order to help increase the knowledgeable use of standardized test results, the author discusses a number of misconceptions under ten headings. In addition, some suggestions to help avoid abusing tests and test scores are made.*

WE ARE in a boom period of standardized testing in elementary and secondary schools. Millions of tests are administered each year to pupils at all grade levels—achievement tests, mental ability tests, aptitude tests, and interest inventories, as well as several other types of tests and inventories. Some of these tests are given for college scholarship purposes and some for college admissions purposes. Title V of the National Defense Education Act has stimulated, and in some instances required, additions to testing programs at the secondary level. In general, however, these external influences account for a relatively small percentage of the total standardized testing undertaken by a school system.

There are at least two factors which have had a greater impact upon the amount of testing done in the schools than NDEA or college requirements.

Reprinted from *Michigan Journal of Secondary Education,* 1962, 2, 153–161, by permission of the publisher.

First, there has been and continues to be a natural growth of standardized testing at all grade levels. Second, the rapid growth of the guidance movement has meant a corresponding rapid growth in testing. This latter influence may well be the most influential one operating, for in many schools the testing program is developed by and operated by guidance personnel.

Inauguration or expansion of a testing program is relatively easy. Decisions made one day can be implemented within a week or two. The only time lag is that of the United States mail in delivering orders for tests and getting test materials from the publisher to the school. Machine methods for test scoring have reduced, and in many cases eliminated, objections that a testing program is a burden upon individual teachers. The school budget and allocation of time for testing are the only real problems to face if an administrator or faculty decides to enlarge the testing program. Thus it is relatively easy to test.

The values of standardized testing, however, cannot be dismissed so quickly. Such values are dependent upon two processes: (1) establishment of proper goals of testing and the development of a testing program to meet those goals, and (2) proper use of test results. Both of these processes are essential to the operation of a successful testing program. Most educators feel, and rightfully so, that the major weakness of testing today is in the area of test use.

Many writers have made this point, and most school administrators are acutely aware of the fact that the ultimate criterion for judging the effectiveness of their testing programs is the correct use of test results. Accumulating test scores in cumulative files is not evidence of test use. Correct use depends upon getting test results into the hands of counselors, teachers, administrators, pupils, and parents *and* of being sure that each consumer of these results is made knowledgeable enough to interpret them. In this latter statement—"made knowledgeable enough to interpret them"—lies the key to proper test use.

There are a number of ways that test scores are misused or overused and a number of misconceptions about tests and test scores that are common enough to warrant special attention. While one could think of innumerable examples of specific errors in test interpretation, the purposes of this article is to point up some of the more common mistakes, in order to help increase the knowledgeable use of test results. Ten points have been selected for discussion; others could have been added.

## Categorizing a Pupil at a Specific Level of Achievement or Ability

One of the most common mistakes made by persons unskilled in interpreting test results is the assumption of perfect reliability of a test score. Too often it is assumed that an IQ of 105 represents performance definitely superior to that represented by an IQ of 104 and definitely inferior to that represented by 106. Too often we fail to realize that a test score is best interpreted as a good estimate of the *general level of performance,* and that it will vary from test to test and from time to time. Test users must accept the concept of variability of test scores over time and over tests. The assessment of human traits and abilities is not at the same level of accuracy as that found in a physics laboratory. It probably is closer to the level of accuracy found in the predictions of weather, in which temperature predictions are generally within a few degrees of actual temperatures, but in which differences of ten or more degrees are common enough to be remembered vividly by critics.

Another aspect of this assumption of greater accuracy than actually prevails is the use of a single estimate (test score) to predict human performance. It is generally wise to insist on having two or three reading scores, or two or three aptitude scores, before putting much confidence in them. This is a direct result of the unreliability present in all test scores. If a pupil receives percentiles of 35 and of 40 on arithmetic tests given in two different years, one can have greater confidence that his level of achievement in arithmetic is in the average range than if only one of these scores is available.

## Confusion of Norms and Standards

Norms are test scores which tell us the level of performance attained by an average or typical group of pupils. Standards represent human judgments of the level of performance that "should be" attained by a group of pupils. A test user should *not* assume that "typical" pupil performance is automatically the "proper" level of performance for pupils in a particular school system. It is reasonable, of course, to assume that pupils in many school systems will tend to perform at a level close to the level of test norms. In others, however, it is reasonable to assume that pupils will perform at a higher level or at a lower level.

One occasionally finds a test user who completely fails to grasp the conception of what a test norm is. Since a test norm represents "typical" performance, then, of necessity, *half* of the pupils in a typical or average group will have scores at or below the average score. If a teacher of a typical group of pupils finds that 40 per cent of his pupils are below grade level in reading, he is to be congratulated. In the norm group for whatever test is being used, 50 per cent of all the pupils were at or below grade level. The assumption that all pupils in a class should be *at* grade level is patently impossible, unless one knows that the poorest achieving pupil in one's class is in the top 50 per cent of all pupils his own age or grade.

## Assumption That Test Scores Predict Success or Failure for Individual Pupils

One way that test results often are overused is the assumption that a particular score or series of scores does in fact predict success or failure with unfailing accuracy. It is well established that students who succeed in colleges of engineering generally make high scores on numerical ability tests. Yet it is not correct to conclude from such data that Johnny, with a 50th percentile rank on a test of numerical ability, will not succeed in an engineering course. It is correct to conclude that of every one hundred students with numerical ability scores the same as Johnny's only a small percentage will succeed in an engineering curriculum. The test score does provide information of a probability type; it enables a student or parent or counselor to know the odds for success or failure. It is a well-known fact that long shots occasionally win the Kentucky Derby, but year in, year out, the favorites generally win.

It is not unusual for two counselors to look at the same test scores for an individual pupil and to come to somewhat different conclusions. For this reason, it is well to face the fact that while test scores do provide information that can be helpful in decision making, the decisions for courses of action are made by human beings, not by the scores.

Added to this overuse of test scores is the failure of some people to utilize all pertinent data available about a student when test scores are known. To allow test scores to outweigh all other judgmental data is a misuse of these scores; to ignore test scores in favor of other judgmental data also is a misuse of these scores.

## Determination of Vocational Goals

"Mary's scores from a clerical speed and accuracy test and from a spelling test are only average. Therefore, Mary should not consider secretarial work as a career possibility." Or, "Since Jim's interest profile shows high scores in 'Scientific' and 'Social Service' he should elect a premed course in college." How often can vocational counseling be summed up in just such simple statements? It is so easy to make the jump from test score to occupation, and it seems so logical that this type of interpretation should be accurate. Unfortunately, the predictive validity of test scores in high school for success in specific occupations is not good enough to permit such interpretations. Most evidence of the predictive efficiency of test scores relates those scores to academic curricula. We can say with a fair degree of accuracy that certain patterns of test scores predict fairly well in different curricula. That is the type of validity data that is generally available.

The use of test scores in vocational counseling should tend to open doors of possible occupations rather than close them. Again, presenting the relationship between test scores and occupational areas on a probability basis can be helpful, and is certainly more accurate than making the assumption that certain test scores assure success in one field and failure in another.

## Assumption That Intelligence and Achievement Are Separate and Distinct

Here are two sample questions from standardized tests:

1. Extraneous
   a. extra b. foreign c. transparent d. noisy
2. Make *indelible* means
   a. indistinct b. permanent c. purple d. identical

Both are vocabulary items. One of them is taken from a widely used intelligence test (California Test of Mental Maturity) and the other from a widely used achievement battery (Iowa Tests of Educational Development). Vocabulary items measure the learned meanings of words; vocabulary items are our best single measure of general intelligence or scholastic aptitude. Arithmetic items and general information items are also found in both achievement tests and intelligence tests. It is true, of course, that some items suitable for an intelligence test (number series, verbal analogies) are not good measures of achievement. It also is true that many direct measures of achievement (capitalization, punctuation, spelling) are not good measures of intellectual potential.

There is considerable overlap between standardized tests of achievement and standardized intelligence tests. One of the important differences between the two is the way the results are used. When analyzing achievement test scores one is generally considering past performance, what has been accomplished. When analyzing intelligence test scores one is generally looking forward to the future, predicting performance.

It is well to keep in mind the fact that intelligence is inferred from achievement. We have no direct measures of intelligence completely divorced from achievement.

### Assumption That Interests and Aptitudes Are Synonymous

Probably few users of standardized tests would acknowledge a belief that interests and aptitudes are the same thing. Webster defines the terms in clearly different domains. Yet how many users of standardized interest inventories can truthfully say that they have never made the jump from a high percentile score in "Persuasive" to the suggestion that Bill probably could succeed in sales activities? To say that Bill seems to be interested in many of the same things that are of interest to people who work in occupations that require influencing other people may be accurate. But to say that Bill will likely succeed in one of these occupations is to make the unwarranted jump from interest to aptitude.

There is evidence that interests and aptitudes are correlated, but not at a level that allows us to predict one from the other with a high degree of accuracy. This is not to say that interest inventories are useless, but their use might well center on their motivational attributes, on their power to stimulate pupil concern over long range planning.

### Misconception of the Meanings of Certain Types of Derived Scores

Students of education have been and are continuing to be taught that an intelligence quotient is obtained by dividing mental age by chronological age, and that mental age is determined by the test performance of students at different age levels. Yet, as a matter of fact, very few of the IQ's to be found in the cumulative folders of elementary and secondary schools today are quotient scores at all. IQ's are standard scores, just as are z scores, T scores, stanines, College Board scores, and others for almost every widely used intelligence test.

It is true that IQ's originally were quotient scores. But, primarily for statistical reasons, the deviation IQ was developed some years ago and has since met with almost universal adoption. Even the Stanford-Binet test switched to a deviation IQ in 1959. The change from a quotient score to a standard score has not necessitated any drastic change in interpretation. Yet it seems to the writer that test users would be well advised to stop paying lip service to a type of score that no longer exists, and to become familiar with standard scores, the type of scores actually being used with our intelligence tests.

The grade placement or grade equivalent score is another type of derived score that is frequently misinterpreted or overused. All too often it is assumed that a grade placement score is an indication of the grade to which a pupil should be assigned. It does not provide that type of information; it simply tells a user whether a pupil is doing high, average, or low quality work. A percentile rank also provides the same assessment of level of work, yet avoids the danger of overinterpretation. If one wishes to compare a pupil's achievement on two different tests in an achievement battery (e.g., reading level versus arithmetic level), a grade equivalent score may lead one to an important misinterpretation. Because of the variability (standard deviation) of grade placement scores from test to test it is possible for a sixth grade pupil to be at the 90th percentile in both reading and arithmetic, yet receive grade placement scores of 8.8 in reading and 8.0 in arithmetic.[1] If a teacher sees only the grade placement scores of 8.8 and 8.0 he may assume superiority in reading, whereas the two scores represent equivalent performance. For test-by-test comparisons in elementary level achievement batteries percentile scores should be used.

## Grading or Promoting Pupils

Standardized achievement tests are designed with certain purposes in mind. In general, test authors attempt to identify those skills and understandings that are common to most educational programs. They look for the common denominators; they make no attempt to cover those unique aspects of content that a particular school system may incorporate in its curricular offering. They cannot attempt to reflect a particular teacher's goals for his own pupils. Thus, while achievement test results represent very useful assessment of certain skills and understandings that are common to many classrooms, they should not be used to replace a teacher's own assessment devices.

In many schools standardized achievement tests are given toward the beginning of the school year. They are used to look ahead rather than to look back, to diagnose rather than to evaluate or grade. In those schools that use standardized achievement tests at the end of the year, it may be interesting for a teacher to compare the results with his own judgments. It is not wise for the test results to be used to replace his judgments, in either grading or promotion.

---

[1] Iowa Tests of Basic Skills, end-of-year percentile norms for sixth grade pupils.

## Judging Effectiveness of Instruction

Just as standardized tests are not designed to be used for grading pupils, they are not designed to be used for grading teachers. Many of the outcomes of classroom instruction cannot be programmed in standardized tests. Those that can be programmed in tests may not be meaningful because of different emphases, different content, and different grade placement in a particular school.

Of special concern is the attitude engendered in teachers in a school attempting to assess instruction through achievement tests. When test results are used to judge teachers, teachers soon learn to teach for the tests.

It is interesting to note that in some instances teachers even feel compelled to "teach for" ability tests. They somehow feel that it isn't respectable to turn in a set of IQ scores for filing in a cumulative record unless all or almost all of them are at least 100. Such a feeling, of course, is based on a misconception of the meaning of intelligence. A teacher may be cutting his own throat with such high scores, for if his pupils all are above average in ability they may be expected to show equally high achievement levels.

## Comparing Results from Different Tests

There is a very natural tendency for test users to assume that a language usage score from one test is directly comparable to a language usage score from another test, that an IQ from one test means the same thing as an IQ from another. When making such assumptions one tends to forget two very important characteristics of standardized tests:

1. Test authors do not build their tests on the same specifications, following the same blueprints. Each one develops his own specifications for test construction. There usually is considerable overlap between the plans for a language usage test developed by one author and the plans developed by another. However, there is never a complete overlap. Scores from two tests measuring the same attribute vary to a certain extent because the test designs vary.

2. The norms for different tests are based upon different groups of pupils. Each test author aims at securing a truly random population of pupils for use in standardizing his test. Each author falls somewhat short of his goal. While it is correct to assume that test norms for two different arithmetic tests are based on groups with considerable overlap in achievement, it is not correct to assume 100-per-cent overlap.

Thus, two IQ's derived from two different intelligence tests are not exactly equivalent. It has been demonstrated that IQ's can vary as much as 5 or 10 points between different tests for no other reason than that they are different tests.

Sometimes one hears this objection: "But how can IQ's be different from different tests? I thought that all good intelligence tests correlated well with each other." It is true that the correlation between different inelligence tests are generally sizable, and many times are almost as high as the reliability of the separate tests. Such correlations do *not* guarantee comparability of norms. Such correlations simply say that pupils taking the different tests will tend to get scores putting them in the same relative rank order but not with the same scores. For example, suppose one were to take a set of IQ's (or any other test score) and add 50 points to each score. The correlation between the original IQ's and the new scores would be perfect, yet the two sets of scores would be 50 points apart.

As was mentioned earlier, we are in the middle of a boom period of testing. If the users of tests do a good job of interpreting the results for the improvement of our understanding of boys and girls and for the improvement of instruction, the boom will level off on a satisfactory plateau of test use. If the users of test results fall into the various misuses, overuses, and misconceptions that are possible, the boom will most certainly be followed by a "bust."

It is the thesis of this article that the consumers of test scores must be thoroughly conversant with proper methods of test use and must studiously avoid misuses, overuses, and misconceptions.

"A little knowledge is a dangerous thing."

# 47

# The Assignment of School Marks

## — *Marvin D. Glock and Jason Millman*

*The authors discuss the use of test results and other assessment procedures in assigning school grades. The teacher is viewed in this connection as having three responsibilities: stating objectives, providing experiences to help students attain those objectives, and determining the extent to which they have been achieved. A basic concept to the present article is that direct conversion of test scores into grades is meaningless. Measures of achievement provide only relative evaluations of a student's progress and not an absolute index of his mastery of course content. Various means of converting raw scores into standard scores are presented.*

A TEACHER has a three-fold responsibility: he must decide what he wants to accomplish with a class (state his objectives); he must provide the experiences that will help his pupils attain those objectives; and he must determine whether the objectives have been achieved.

In determining achievement, the teacher makes a value judgment based on measures such as tests, grades on projects, and ratings on laboratory performances. Once the measures have been obtained, it is necessary to convert them into a final grade. The purpose of this bulletin is to discuss the conversion procedure.

Reprinted from *Cornell Miscellaneous Bulletin 44*, New York State College of Agriculture, Cornell University, 1963, by permission of the publisher.

### A Basic Concept

When a student correctly answers seven of 10 questions about the Civil War, it does not mean that he knows 70 percent of the facts about the War. If the questions were easier or harder, the student could get 0 percent or 100 percent on the test. To illustrate, consider tests A and B which were given to the same group of students. No one did better on test A than on test B; several students received 100 percent on test B and 0 percent on test A. Our judgment of how much a student knows about world geography and maps could vary greatly depending upon which test he took.

| TEST A | TEST B |
|---|---|
| 1. Give a rank of 1 to the continent with the largest area; a rank of 3 to the continent with the smallest area.<br>_____Antarctica<br>_____Europe<br>_____South America | 1. Give a rank of 1 to the continent with the largest area; a rank of 3 to the continent with the smallest area.<br>_____Africa<br>_____Asia<br>_____Australia |
| 2. What is the name of a map projection that represents a "true picture" projection of any given half of the earth? ____________ | 2. Which of the following types of map projections has the longitude and latitude lines at right angles to each other?<br>_____a) Conic projection<br>_____b) Mercator projection<br>_____c) Polar projection<br>_____d) Sinusoidal projection |
| 3. Lake Geneva is really a part of the<br>_____a) Rhine River<br>_____b) Rhone River<br>_____c) Seine River<br>_____d) Volga River | 3. A famous river valley in China is the<br>_____a) Amazon<br>_____b) Po<br>_____c) Rhine<br>_____d) Yangtze |

The actual score on a test does not tell us how well a student has mastered the material for which we are testing. To convert test scores directly to grades is meaningless, but test scores do permit us to make *relative judgments* about the students who took the tests. To arrive at a final grade, we need some idea of how Henry's test scores and other measures of performance compare with those for other students being graded. To provide a basis for making comparisons, it is necessary to combine the component measures into a composite value or score.

## Determining a Composite Score

When combining scores on informal quizzes, reports, class participation, final examinations, and so forth, one must decide what weight each component shall be given. There is no consensus on the exact emphasis or weight various measures should have, but it is agreed that measurements reflecting pupils' progress in the more important course objectives should be given greater weight than those reflecting progress toward less important objectives. For example, if a French teacher feels that the ability to carry on a conversation in French is more important than knowledge of French history, her measures of students' conversational skill should have more weight than a test score in French history.

## A Necessary Precaution

A precaution must be taken in determining a composite mark, however, no matter what weight is chosen. Applying a weight to test scores is insufficient because of distortion from inequalities in score variability. Variability is the degree of spread or dispersion of a set of test scores along the score range. The more variable component score will receive greater weight than desired. This can be illustrated by a simple example in which a teacher wishes to weight two test scores equally. Bill received the highest score in the class on test 1 and the lowest score in the class on test 2; Mary had the lowest score in the class on test 1 and the highest score in the class on test 2. Their scores and the tests' range (variability) are shown below:

|  | Test 1 | Test 2 | Composite Score |
|---|---|---|---|
| Bill | 90 | 50 | 140 |
| Mary | 70 | 80 | 150 |
| Test Range | 20 | 30 |  |

Because both Bill and Mary had one top score and one bottom score, and because each test was to be weighted equally, each student should receive the *same* composite score. Mary's composite score was higher because her better performance was on the test with the greater variability.

Making the average of each test the same is not the solution because test 1, which has the higher average, exerts less influence. Changing the test scores to percentages is not the solution either, because these scores could as well have been percentage scores. The scores must be converted in such a way that each test has the same variability. Because the original scores provide us with relative judgments, the converted scores should not disturb the ranking of the students.

### Stanines

To control the variability of test scores, test makers frequently change the original test scores to standard scores. For example, College Board test scores are reported by a type of standard score. Conversion of the obtained scores of all the component measures to any one of the many types of standard scores will equalize the variability of the several component scores. Stanines, a type of standard score rapidly growing in popularity, have been selected for illustration because no calculation is necessary to convert obtained scores to stanines.

Stanines are a STAndard NINE-point scale in which 1 represents the lowest scale value and 9 the highest scale value. To convert a set of obtained measures into stanines, it is necessary to rank the measures. Table 1 indicates how many students should receive each stanine value. For example, if 33 measures were being ranked, one student would receive a stanine of 1; two students a stanine of 2; four students a stanine of 3; six students a stanine of 4, and so forth.

To illustrate, a teacher has ranked 15 papers with scores as follows:

30 28 27 27 25 25 25 23 23 21 20 20 7 5 0

The teacher might next place a slanted line between the scores to separate them into stanine groups according to the breakdown in the preceding table.

30/ 28/ 27 27/ 25 25/ 25 23 23/ 21 20/ 20 7/ 5/ 0

The two top scores will receive stanines of 9 and 8 respectively and the two scores tied at 27 will, of course, both receive a 7. At this point the teacher needs to be careful to give the three scores tied at 25 the same stanine value. Because the majority of the 25's are in the 6th stanine level, assigning all a 6 results in the least distortion. The next problem involves the two scores tied at 20. Because the 20 is more nearly a 21 than a 7, converting both 20's to stanine of 4 is the least distorting procedure.

There are no hard and fast rules for assigning stanine values to tied scores that span more than one stanine level. Common sense will be your best guide, and a shift of one stanine level makes little difference. The original scores, with their stanine equivalents, are shown below:

30/ 28/ 27 27/ 25 25/ 25 23 23/ 21 20/ 20 7/ 5/ 0

9 8 7 7 6 6 6 5 5 4 4 4 3 2 1

TABLE 1. Number of Students Receiving Each Normalized Stanine Score

| Number Being Measured | Stanine Score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| | Number of students receiving score | | | | | | | | |
| 15 | 1 | 1 | 2 | 2 | 3 | 2 | 2 | 1 | 1 |
| 16 | 1 | 1 | 2 | 3 | 2 | 3 | 2 | 1 | 1 |
| 17 | 1 | 1 | 2 | 3 | 3 | 3 | 2 | 1 | 1 |
| 18 | 1 | 1 | 2 | 3 | 4 | 3 | 2 | 1 | 1 |
| 19 | 1 | 1 | 2 | 3 | 5 | 3 | 2 | 1 | 1 |
| 20 | 1 | 1 | 2 | 4 | 4 | 4 | 2 | 1 | 1 |
| 21 | 1 | 1 | 2 | 4 | 5 | 4 | 2 | 1 | 1 |
| 22 | 1 | 2 | 2 | 4 | 4 | 4 | 2 | 2 | 1 |
| 23 | 1 | 2 | 2 | 4 | 5 | 4 | 2 | 2 | 1 |
| 24 | 1 | 2 | 3 | 4 | 4 | 4 | 3 | 2 | 1 |
| 25 | 1 | 2 | 3 | 4 | 5 | 4 | 3 | 2 | 1 |
| 26 | 1 | 2 | 3 | 4 | 6 | 4 | 3 | 2 | 1 |
| 27 | 1 | 2 | 3 | 5 | 5 | 5 | 3 | 2 | 1 |
| 28 | 1 | 2 | 3 | 5 | 6 | 5 | 3 | 2 | 1 |
| 29 | 1 | 2 | 4 | 5 | 5 | 5 | 4 | 2 | 1 |
| 30 | 1 | 2 | 4 | 5 | 6 | 5 | 4 | 2 | 1 |
| 31 | 1 | 2 | 4 | 5 | 7 | 5 | 4 | 2 | 1 |
| 32 | 1 | 2 | 4 | 6 | 6 | 6 | 4 | 2 | 1 |
| 33 | 1 | 2 | 4 | 6 | 7 | 6 | 4 | 2 | 1 |
| 34 | 1 | 3 | 4 | 6 | 6 | 6 | 4 | 3 | 1 |
| 35 | 1 | 3 | 4 | 6 | 7 | 6 | 4 | 3 | 1 |
| 36 | 1 | 3 | 4 | 6 | 8 | 6 | 4 | 3 | 1 |
| 37 | 2 | 3 | 4 | 6 | 7 | 6 | 4 | 3 | 2 |
| 38 | 1 | 3 | 5 | 6 | 8 | 6 | 5 | 3 | 1 |
| 39 | 1 | 3 | 5 | 7 | 7 | 7 | 5 | 3 | 1 |
| 40 | 1 | 3 | 5 | 7 | 8 | 7 | 5 | 3 | 1 |
| 41 | 1 | 3 | 5 | 7 | 9 | 7 | 5 | 3 | 1 |
| 42 | 2 | 3 | 5 | 7 | 8 | 7 | 5 | 3 | 2 |
| 43 | 2 | 3 | 5 | 7 | 9 | 7 | 5 | 3 | 2 |
| 44 | 2 | 3 | 5 | 8 | 8 | 8 | 5 | 3 | 2 |
| 45 | 2 | 3 | 5 | 8 | 9 | 8 | 5 | 3 | 2 |
| 46 | 2 | 3 | 5 | 8 | 10 | 8 | 5 | 3 | 2 |
| 47 | 2 | 3 | 6 | 8 | 9 | 8 | 6 | 3 | 2 |
| 48 | 2 | 3 | 6 | 8 | 10 | 8 | 6 | 3 | 2 |
| 49 | 2 | 4 | 6 | 8 | 9 | 8 | 6 | 4 | 2 |
| 50 | 2 | 3 | 6 | 9 | 10 | 9 | 6 | 3 | 2 |
| 51 | 2 | 3 | 6 | 9 | 11 | 9 | 6 | 3 | 2 |
| 52 | 2 | 4 | 6 | 9 | 10 | 9 | 6 | 4 | 2 |
| 53 | 2 | 4 | 6 | 9 | 11 | 9 | 6 | 4 | 2 |
| 54 | 2 | 4 | 7 | 9 | 10 | 9 | 7 | 4 | 2 |
| 55 | 2 | 4 | 7 | 9 | 11 | 9 | 7 | 4 | 2 |
| 56 | 2 | 4 | 7 | 9 | 12 | 9 | 7 | 4 | 2 |
| 57 | 2 | 4 | 7 | 10 | 11 | 10 | 7 | 4 | 2 |
| 58 | 2 | 4 | 7 | 10 | 12 | 10 | 7 | 4 | 2 |
| 59 | 3 | 4 | 7 | 10 | 11 | 10 | 7 | 4 | 3 |
| 60 | 3 | 4 | 7 | 10 | 12 | 10 | 7 | 4 | 3 |
| 61 | 3 | 4 | 7 | 10 | 13 | 10 | 7 | 4 | 3 |
| 62 | 3 | 4 | 7 | 11 | 12 | 11 | 7 | 4 | 3 |
| 63 | 3 | 4 | 7 | 11 | 13 | 11 | 7 | 4 | 3 |
| 64 | 3 | 4 | 8 | 11 | 12 | 11 | 8 | 4 | 3 |
| 65 | 3 | 4 | 8 | 11 | 13 | 11 | 8 | 4 | 3 |
| 66 | 3 | 4 | 8 | 11 | 14 | 11 | 8 | 4 | 3 |
| 67 | 3 | 5 | 8 | 11 | 13 | 11 | 8 | 5 | 3 |
| 68 | 3 | 5 | 8 | 11 | 14 | 11 | 8 | 5 | 3 |
| 69 | 3 | 5 | 8 | 12 | 13 | 12 | 8 | 5 | 3 |
| 70 | 3 | 5 | 8 | 12 | 14 | 12 | 8 | 5 | 3 |
| 71 | 3 | 5 | 8 | 12 | 15 | 12 | 8 | 5 | 3 |
| 72 | 3 | 5 | 9 | 12 | 14 | 12 | 9 | 5 | 3 |
| 73 | 3 | 5 | 9 | 12 | 15 | 12 | 9 | 5 | 3 |
| 74 | 3 | 5 | 9 | 13 | 14 | 13 | 9 | 5 | 3 |

NOTE: This procedure should probably not be used for classes of less than 15. If more than 74 students are being measured, the distribution of stanines may be determined by adding together two or more rows. For example, if 77 students are being measured, add together such rows as "37" and "40." If this is done, then 2+1 or 3 of the 77 students should receive a stanine of one; 3+3 or 6 students should receive a stanine of two; and so forth.

By converting test scores and other measures into stanines, most students' progress in comparison to others in a class can be easily noted during a year. An exception, of course, is the student who does best (stanine = 9) on one measure and cannot be higher on another. A very low scoring student may actually improve his achievement, but if the other members of the class improve as well he may remain in the bottom stanine level. In addition, differences of less than three stanine levels on two measures probably do not represent statistically significant changes because of the low reliability of most teacher-made tests.

Stanines also provide a basis for comparing intelligence and achievement scores which can be converted into stanines according to the procedure described above. All things being equal, a pupil should be achieving at the stanine level of his scholastic aptitude. We would probably consider Johnny an underachiever if his intelligence score were at the ninth stanine and his achievement at the fifth, although a difference of less than three stanine levels between these two measures is not significant.

### The Composite Average

Because distributions of stanine scores all have the same variability, the desired weight can be assigned to component scores being combined into a composite by applying the following formula:

$$C = \frac{(W_1 x S_1) + (W_2 x S_2) + (W_3 x S_3) + \cdots}{W_1 + W_2 + W_3 + \cdots}$$

C is the composite average; $W_1$ is the weight and $S_1$ is the stanine level on the first component measure; $W_2$ and $S_2$ are the weight and stanine level on the second component measure, and so forth.

To illustrate, suppose that a student has the following stanine scores on four component measures to be weighted as indicated:

| Composite Measure | Weight | Stanine |
|---|---|---|
| Unit Test #1 | 1 | 3 |
| Unit Test #2 | 1 | 2 |
| Reports | 1 | 4 |
| Final Examination | 2 | 6 |

His composite average would be:

$$C = \frac{(1x3) + (1x2) + (1x4) + (2x6)}{1 + 1 + 1 + 2} = \frac{21}{5} = 4.2$$

If the student were absent for the first unit test, his composite average would be:

$$C = \frac{(1x2) + (1x4) + (2x6)}{1 + 1 + 2} = \frac{18}{4} = 4.5$$

In the earlier example of Bill and Mary, each student received a top and a bottom score on two tests of equal weight. In each case their composite average would be:

$$C = \frac{(1\text{x}9) + (1\text{x}1)}{1 + 1} = \frac{10}{2} = 5.0$$

Although composite averages are not stanines, they do provide a means for ranking students which reflects the desired relative weight for several measures. On the basis of this distribution, the teacher may assign final marks.

## Assigning a Final Grade

Once a composite score is computed for each student, and the composite scores ranked, final grades can be assigned. In absolute terms, a composite score tells little about how much a student knows, but it has meaning in comparison to composite scores of others in a class. Because test scores have no absolute zero, all scores indicate relative judgments. Standardized tests that measure objectives in which the school is interested can give an idea of the degree to which standards are being met. A school that requires a percentage score of 75 to pass need not have higher standards than a school in which 65 percent is passing. Tests can be made easier or harder at the whim of the teacher. Teachers who convert percentage scores directly into school marks are constantly adjusting by having make-up tests, raising marks by some mysterious formula, making the next test easier, altering their grading standards, or some other device. Whether they wish to admit it or not, many teachers decide in advance what the frequency of each grade they are going to give will be, and adjust their composite measures accordingly.

### How Many A's?

The distribution of final grades involves an arbitrary decision. It may depend upon school policy, the ability or vocational plans of the students, the presence of gaps in the distribution of composite scores, and so forth. Oddly enough, most schools do not consider a C an average grade. Most schools give more A's and B's than D's or E's, so the typical grade is nearer a B than a C.

### Giving Grades Meaning

Regardless of the marking system (number grades, letter grades, graphs) it is important that school grades present a clear and uniform meaning to all who use them. Because grades should reflect a student's achievement in a particular subject, effort, neatness, punctuality, and other nonachievement component measures should not enter into the final calculation. If it is desirable to indicate these traits, a separate grade should be given.

School personnel should agree on the criteria for determining the percentage of pupils in a class to be assigned a given grade. Classes in the same subject with students of similar ability should receive the same grade distribution. It may be desirable to give a higher percentage of students a top grade in certain subjects (physical education), in honors sections, and at the elementary school level, but these decisions should be made by the school personnel.

To illustrate how grades may be assigned unfairly, consider Tom, Dick, and Harry who each have an achievement in history comparable to that of the average college-bound student. Tom is in a college-bound class where the teacher gives only 40 percent A's and B's, and Tom gets a C. Dick is also in a college-bound class, but his teacher gives 60 percent A's and B's, and he receives a B in history. Harry, equal in achievement to Tom and Dick, is in a history class for students in a general curriculum. In this class, Harry is one of the best students and he receives an A. Despite their equal achievement in history, Tom, Dick, and Harry receive grades of C, B, and A respectively because they are in different classes.

## Adjusting for Differences in Ability of Classes in the Same Subject

One way to rectify the inequality in the case of Tom, Dick, and Harry is to obtain some common and relevant measure by which the three history classes could be compared. Scores from a scholastic aptitude test, a standardized history test, or a common examination could be used. To see how this would work, suppose it is decided that 18 percent of the 150 students taking history shall receive a grade of A. When the students are ranked according to the common measure, suppose that 10 of the top 27 (18 percent x 150 = 27) scores are made by students in an honors section. Consequently, the teacher of the honors class will give A's to the 10 students with the highest composition averages, although they need not be the 10 who scored highest on the common measure. It is possible that a class would have no pupils who received one of the top 27 scores. Therefore, no A's would be given in that class.

Grades for other subjects would be distributed in a similar manner. The principle would be the same, although the percentages of each grade may differ.

## Summary

Measures of achievement can provide only a relative judgment about a student. They do not represent an absolute percentage of mastery in a course. A defensible procedure in assigning grades is to determine a pupil's composite score in a subject, giving careful consideration to weight. To prevent a bias in weighting, it is helpful to convert raw scores into standard scores such as stanines.

After composite scores are determined and ranked, grades may be assigned. Assignment of grades involves an arbitrary decision about the percentage of each grade to be awarded. Percentages will differ on the basis of subject matter, ability grouping, and vocational goals. When there is more than one class in a subject area, it is wise to consider a procedure providing for a fair distribution of grades.

## REFERENCES

Ahmann, J. S., & Glock, M. D. *Evaluating pupil growth.* Boston: Allyn and Bacon, 1958.

Durost, W. N. The characteristics, use and computation of stanines. *Test service notebook* #23. New York: Harcourt, Brace & World, Inc., 1961.

Johnson, Mauritz. "Solving 'the Mess in Marks'," *New York State Educ.*, 1961, **49**, 12–13, 30.

Thorndike, R. L., & Hagen, Elizabeth. *Measurement and evaluation in psychology and education.* New York: John Wiley, 1961. Chapter 17.

# 48

# The Effect of Promotion Policy on Academic Achievement — *G. T. Kowitz and C. M. Armstrong*

*A longitudinal study of the effect of promotion policy on academic achievement, by use of "allometric charts," is reported in this article. Records of students in two school districts that had contrasting promotion policies were studied and compared. It was discovered that students' academic achievement does respond to the promotion policy of the school. However, there was more change among students who were promoted than among students who were retained. The authors urge consideration of the implications to mental health of a policy based on the threat of nonpromotion.*

DURING THE past several years there has been a growing demand for a tough program of education in the public schools, particularly in the elementary grades where a concern for the whole child is often stressed. But few educators have bothered to distinguish between a program that is academically tough and a program that is personally tough. Many prefer to rely on the archaic maxim that to be good, medicine must taste bad.

Actually, there is reason to believe that an academically tough program, properly administered, far from being unpleasant for the pupil, would be

Reprinted from *Elementary School Journal* 1961, **61,** 435–443, by permission of the senior author and the University of Chicago Press.

quite satisfying to him. There is pleasure in true accomplishment even though it calls for strenuous effort and may be fatiguing.

All too often the educational program is judged tough or soft by one test only: the promotion policy of the school. Proponents of a tough program often suggest that the school with the higher rate of pupil failure is the better one.

Many studies of the effect of school promotion policy have been concerned primarily with the pupil's personal-social development, not his academic achievement. The work of Sandin (*1*) and of Morrison and Perry (*2*) indicate that retention of pupils because of their inability to achieve academically can have undesirable effects on their personal-social development.

Furthermore, studies of academic achievement and promotion have been concerned with results for a group of pupils or for all pupils in a school system. Researchers have tended to ignore the effect of promotion policy on the achievement of individual pupils (*3, 4*).

Evidence is lacking that retention increases the pupil's academic proficiency. Could the lack of evidence be in part a result of the research approach used?

Promotion is a matter of school policy, outlined by the administration. Since policy is district-wide, the proper unit of analysis is the district, not the pupil within the district.

To evaluate policy, the academic achievement of pupils in school districts that use non-promotion as a device to insure achievement should be compared with the achievement of pupils in districts that do not use this device. In other words, we must allow for two effects—the effect of the threat of non-promotion on pupils who are promoted and the effect of the fact of non-promotion on pupils who are retained.

Furthermore, since pupils do not respond uniformly to threats and coercion, the effect of a policy of promotion based solely on achievement should be sought in the achievement records of individual pupils, not in average scores for school districts. These records must be extended over enough years so that the ultimate effect of the policy can be discerned. A sound administrative policy should insure that each pupil will achieve within his range of ability. The pupil should not be held to some artificial standard set for all pupils in the district.

While this analysis of the problem seems reasonable, it presents several methodological difficulties. Most obvious is the recommendation for a longitudinal approach to the analysis of data on achievement. Adequate longitudinal data are difficult and expensive to collect; and, in spite of our concern for patterns in child development, only a limited methodology has been developed for handling longitudinal data.

Too often researchers turn to individual case histories, which cannot provide general rules. Or investigators simply tack together a series of latitudinal studies, which lack the continuity of individual growth and thus do not have the power of the longitudinal approach.

Another problem arises from current theories of school administration. In an attempt to meet the changing demands of the local community and the greater society, many schools are no longer holding firmly to a determined course of action. Instead they are resorting to a policy of adaptability. In many cases, this approach results in changes in policy before the effects of a policy are fully understood. The need for continuing policy and the requirements of longitudinal research add tremendously to the obstacles to definitive inquiry.

In the records of the Research Offices of the New York State Education Department, it was possible to locate two school districts that had contrasting promotion policies and a reasonable amount of recorded data. As can be seen from Table 1, these two districts seem to be fairly comparable. Several important questions may, therefore, be explored:

TABLE 1. Comparative Statistics of Two School Systems

| | School A | School B |
|---|---|---|
| Enrolment in Grade 10 | 169 | 118 |
| Number of students in Grade 10 who had failed before Grade 7 | 10 | 31 |
| Per cent of students in Grade 10 who had failed before Grade 7 | 5.92 | 26.27 |
| Socioeconomic index | 3.80 | 3.33 |
| Composite achievement (mean) | 14.44 | 15.02 |
| Lorge-Thorndike intelligence quotient (mean) | 108 | 111 |

1. Can longitudinal records be treated in such a way that their uniqueness can contribute to the evaluation of administrative policy?
2. Is a child's pattern of academic growth immutable, or can it be shown to respond to administrative policy?
3. In the schools studied, can it be demonstrated that the promotion policy had the effect the administration intended?
4. Did the policy increase the achievement of the student who did not meet the established standards?
5. Did the academic growth of the individual pupil and the level of achievement for the district respond as a system controlled and regulated by administrative policy?

If it can be shown that there is an orderly response to policy in the achievement of the student body—a response that is not masked by differences in individual responses—it will be accepted that a longitudinal analysis of achievement records can be used to describe the influence of administrative policy.

It is commonly believed that more boys than girls fail in school. Sometimes the superiority imputed to girls has been ascribed to differences in developmental patterns. If girls are superior for developmental reasons, retention as a policy cannot be justified, since it would be impossible for the less developed boy to meet the standards. The difference in the failure rate of boys and girls is sometimes attributed to the prejudice of teachers who favor the more conforming girls. A policy of failure based on prejudice against males cannot be justified.

A sex-linked difference did not appear in School B, which had a failure rate of 26.27 per cent. Among the girls, the rate was 25.0 per cent, while among the boys it was 27.6 per cent. Since the difference between these rates was not significant, it would seem that in this school promotion and retention of pupils was probably based on achievement rather than on sex bias or inadequate development.

In School B a latitudinal study was made of the difference in achievement between groups of pupils who had been retained before Grade 7 and a proportional random sample of pupils who had not been retained. When the data for achievement in reading and arithmetic were treated by analysis of variance, significant differences were found between groups that had failed and groups that had not.

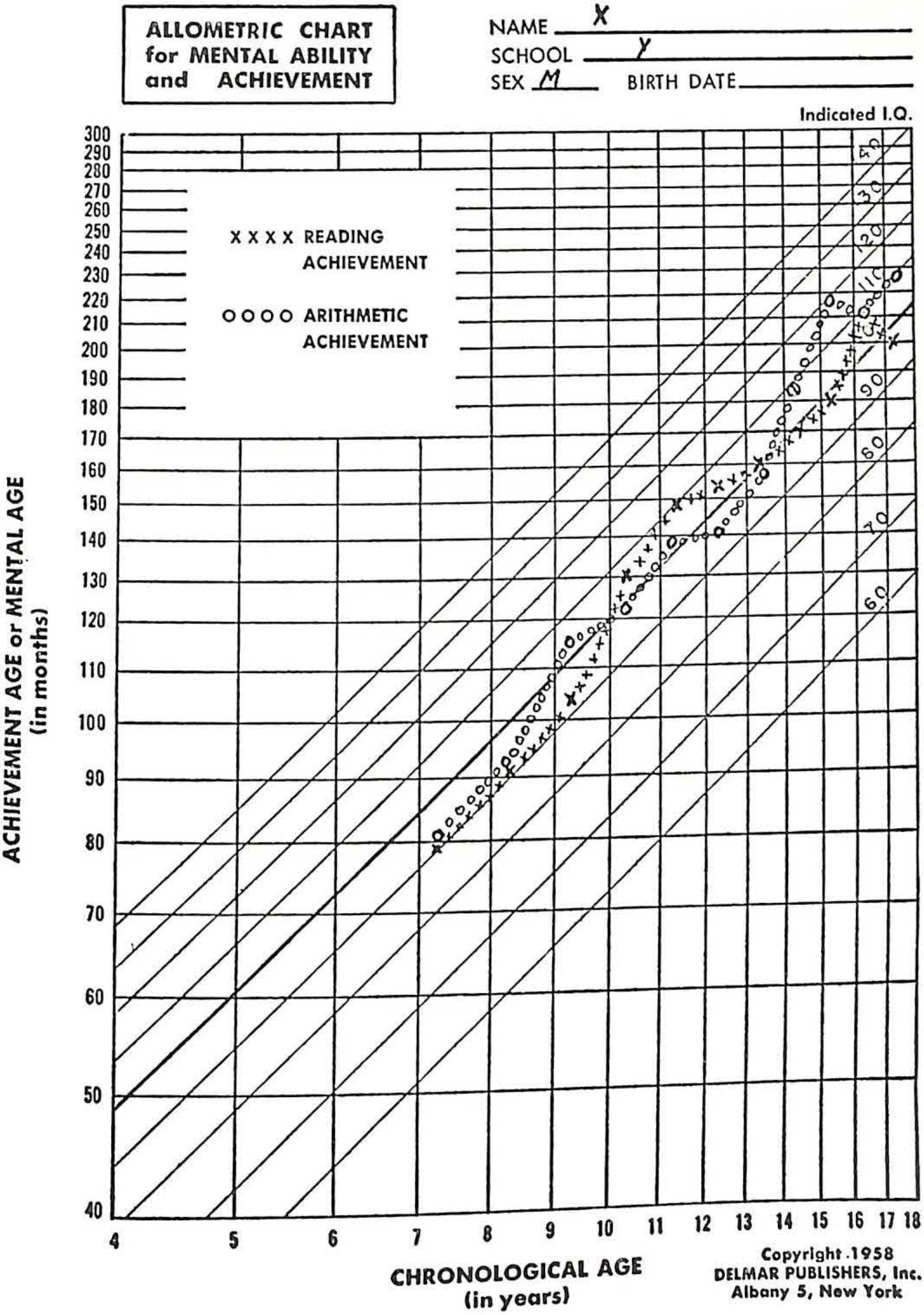
In reading and arithmetic, the achievement of the group that had not been retained was well above the achievement of the group that had been retained. The fact that those who had not made normal progress continued to form a separate group that showed less achievement, even after retention, suggests that their extra year had not eliminated the difference in their achievement. This finding is especially notable since as a result of their retention they were being compared with a group that was a year younger and less experienced.

Thus, even when pupil retention was used as an administrative device to encourage achievement and to limit the range of pupil achievement that the teacher must face in the classroom, pupils who were retained formed a distinct group at a lower level of achievement.

A group treatment of longitudinal data was achieved by using the Allometric Chart as a standard or model (5).

On this chart, which combines the principles of Huxley's law of relative growth (6) and Olson's concept of growth age (7), test scores form a diagonal pattern across the graph.

To help interpret the pattern, diagonal lines, based on mental age, are used to form "channels of development." The boundaries of each channel are determined by lines that show the relationship between mental age and



**ALLOMETRIC CHART for MENTAL ABILITY and ACHIEVEMENT**

NAME X

SCHOOL Y

SEX M BIRTH DATE

Copyright 1958
DELMAR PUBLISHERS, Inc.
Albany 5, New York

FIG. 1. *Achievement record of a boy showing a well-regulated pattern.*

chronological age. Thus it is possible to think of each channel as being ten IQ points wide. Patterns of growth can be said to be described by the number of IQ points encompassed. Research has shown that most patterns of development can be encompassed by three channels, or thirty IQ points (8). This span corresponds closely to the standard error of the classical intelligence test.

Theoretically, the pattern of achievement lies parallel to the channel limits. However, some children show patterns of achievement that move across channels rather than along them. The movement provides a means of classifying children by their pattern of achievement. The achievement of most children, like the boy whose record is presented in Figure 1, shows a pattern of regular progress up the channels. Patterns that cross channel limits can be further classified by the direction of movement—a gain, or an upward movement, and a loss, or a downward movement—and by the number of channels that the pattern crosses.

Besides presenting a picture of the academic growth of individual pupils, the Allometric Chart provides a basis for the statistical treatment of longitudinal data. While the school administrator must always be aware of individual pupils and their particular problems, his policy decisions must be based on significant characteristics of the group. Charts of individual development can be grouped according to the pattern of achievement shown by the child. Five such patterns, or classes, were posited:

Class 1.—Charts that showed a gain of two channels
Class 2.—Charts that showed a gain of one channel
Class 3.—Charts that showed no change of channel
Class 4.—Charts that showed a loss of one channel
Class 5.—Charts that showed a loss of two channels

The response to policy could be observed by noting the proportion, or the per cent, of pupils who evidenced a certain class of development. The effect of the policy could be assessed by comparing the per cent of pupils in each class of development under the contrasting policies of promotion in the two schools.

For example, by comparing the per cent of pupils in each school who showed an increase of channel in their pattern of achievement, the accelerating effect of promotion policy could be evaluated. The significance of the observed differences between the per cent of such pupils in the two districts could be determined from calculations of the standard error of the per cent or proportion (9).

A comparison of the classes of development in the schools indicated that there was indeed a systematic response. No significant differences in pattern were found between boys and girls. The tabulations indicated that Class 5, the group that showed a loss of two channels, was superfluous. Records that would fall into this classification were extremely rare. The comparison of the per cent of pupils in each school district who showed a certain pattern of development suggested that to some extent the pattern was a response to administrative policy. It must be recognized that the differences found might be more correctly ascribed to the effect of the total administrative climate of the schools, of which the rate of pupil retention was but one aspect. However, it is known that in the school with the high rate of retention, the purpose of the policy was to "maintain standards of achievement."

In comparing the per cent of pupils in each channel class, the difference between schools was evaluated in terms of the index of significance. The index is defined as the difference between the per cents measured in standard errors of the difference of the per cents. Substantially, this index is analogous to the *t* test for significance between means. The results are presented in Tables 2 and 3.

Table 2. Comparison of Patterns of Reading Achievement in Two School Systems

| | School A: Low Rate of Retention | | School B: High Rate of Retention | | |
|---|---|---|---|---|---|
| Channel Class | Number of pupils | Per cent | Number of pupils | Per cent | Index of Significance |
| Gain two channels | 4 | 2 | 30 | 25 | 2.982° |
| Gain one channel | 45 | 27 | 37 | 31 | 0.724 |
| No change | 78 | 46 | 36 | 31 | 2.476† |
| Lose one channel | 36 | 21 | 15 | 13 | 1.733 |
| Lose two channels | 6 | 4 | — | — | — |
| Total | 169 | 100 | 118 | 100 | |

° Indicates a difference that is significant beyond the 1 per cent level of confidence.
† Indicates a difference that is significant beyond the 5 per cent level of confidence.

Table 3. Comparison of Patterns of Arithmetic Achievement in Two School Systems

| | School A: Low Rate of Retention | | School B: High Rate of Retention | | |
|---|---|---|---|---|---|
| Channel Class | Number of pupils | Per cent | Number of pupils | Per cent | Index of Significance |
| Gain two channels | 9 | 5 | 25 | 21 | 3.40° |
| Gain one channel | 68 | 40 | 49 | 42 | 0.25 |
| No change | 82 | 49 | 30 | 25 | 3.62° |
| Lose one channel | 9 | 5 | 14 | 12 | 1.55 |
| Lose two channels | 1 | 1 | — | — | — |
| Total | 169 | 100 | 118 | 100 | |

° Indicates a difference that is significant beyond the 1 per cent level of confidence.

From Table 2 it can be seen that School B, which had a high rate of failure, had a significantly higher proportion of pupils who showed a gain of two channels in reading. Also, there was a tendency for this school to have a smaller proportion of pupils who did not change channel. Apparently, the administrative attitude caused an increase in channel among pupils who otherwise might have maintained their course. This achievement beyond expectation suggests that these pupils may have been pressured toward overachievement.

In arithmetic achievement, as Table 3 shows, the changes were similar to those found in reading achievement. When retention was used as a penalty for low achievement, a larger proportion of pupils showed a gain, and a smaller proportion showed no change. Again, the suggestion of pressure toward achievement must be recognized in this shift.

These comparisons present interesting material for elementary-school administrators. Apparently a policy of high retention can result in a greater proportion of pupils who show an increasing rate of achievement. If the school were concerned only with achievement as measured by objective tests, this increase would indeed be laudable. However, the school has a much broader responsibility, and the warnings of mental hygienists on the effect that forced overachievement has on the personality development of the child must be recognized.

That significant gains occurred is of interest. There remains the problem of relating these gains to policy on pupil promotion. A further analysis was made of the patterns in the school that had the high rate of failure. The group that had been retained was compared with the group that had not The results for the reading patterns and the arithmetic patterns are presented in Tables 4 and 5, respectively.

TABLE 4. Comparison of Patterns of Reading Achievement of Pupils Who Were Retained and the Total Grade Population in School B

| | Pupils Retained | | Grade Population | | |
|---|---|---|---|---|---|
| Channel Class | Number | Per cent | Number | Per cent | Index of Significance |
| Gain two channels | 2 | 6 | 30 | 25 | 2.97° |
| Gain one channel | 5 | 16 | 37 | 31 | 1.84 |
| No change | 11 | 36 | 36 | 31 | 0.53 |
| Lose one channel | 13 | 42 | 15 | 13 | 3.05° |
| Lose two channels | — | — | — | — | — |
| Total | 31 | 100 | 118 | 100 | |

° Indicates a difference that is significant beyond the 1 per cent level of confidence.

While the analysis of reading patterns (Table 4) shows that the pressure toward achievement caused a true gain in channel for the group as a whole, it also identified the failure group with a tendency to lose channel. It can be concluded that, while retention segregated the failure group, it did little to

TABLE 5. Comparison of Patterns of Arithmetic Achievement of Pupils Who Were Retained and the Total Grade Population in School B

| Channel Class | Pupils Retained |  | Grade Population |  | Index of Significance |
|---|---|---|---|---|---|
|  | Number | Per cent | Number | Per cent |  |
| Gain two channels | 3 | 10 | 18 | 15 | 0.83 |
| Gain one channel | 4 | 13 | 45 | 38 | 3.19[°] |
| No change | 17 | 55 | 48 | 41 | 1.36 |
| Lose one channel | 6 | 19 | 7 | 6 | 1.80 |
| Lose two channels | 1 | 3 | — | — | 1.02 |
| Total | 31 | 100 | 118 | 100 |  |

[°] Indicates a difference that is significant beyond the 1 per cent level of confidence.

help the pupil whose pattern of reading achievement showed a loss of channel.

The analysis of arithmetic patterns (Table 5) roughly paralleled that of reading patterns. While the group of pupils who were retained showed a significantly smaller proportion gaining channel, there was no change in the proportion losing channel.

The school administrator is often faced with the problem of evaluating policy. There is a great danger in basing judgments on tradition, on over-simplified rules of thumb, or on fragmentary evidence. An attempt was made to develop a longitudinal evaluation of promotion policy. Academic achievement apparently responds to school policy. A policy of "achieve or fail" seems to cause more change among pupils who are being promoted than among pupils who are being retained. While there was a trend toward increased achievement in the school with an "achieve or fail" policy, the increase was limited largely to pupils who were in no real danger of being retained. Since they were in no danger of academic failure, the role of the threat of non-promotion in the mental health of the child must be considered.

The analysis of patterns of achievement-test scores on a longitudinal basis appears to be a productive approach to evaluation of administrative policy on promotion. The approach may be useful in evaluating other policies as well.

## REFERENCES

1. Sandin, A. A. Social and emotional adjustments of regularly promoted and nonpromoted Pupils. *Child Develpm. Monogr.*, No. 32. New York: Bureau of Publications, Teachers College, Columbia Univer., 1944.
2. Morrison, I. E., & Perry, I. F. Acceptance of overage children by their classmates. *Elementary Sch. J.*, 1956, **56**, 217–220.

3. Coffield, W. H., & Blommers, P. Effects of non-promotion in educational achievement in the elementary school. *J. Educ. Psychol.*, 1956, 47, 235–250.
4. Hall, W. F., & Demarest, R. Effects on achievement scores of a change in promotional policy. *Elementary Sch. J.*, 1958, 58, 204–207.
5. *An allometric chart for mental ability and achievement.* Albany, N.Y.: Delmar Publishers, 1958.
6. Huxley, J. *Problems in relative growth.* New York: Dial Press, 1932.
7. Olson, W. *Child development.* (2d ed.) Boston: D. C. Heath, 1959.
8. Kowitz, G. T., Armstrong, C. M., & Wohlferd, G. H. Academic characteristics of college bound students. Report to American Educational Research Association, Chicago, 1961.
9. Wallis, W. A., & Roberts, H. V. *Statistics: A new approach.* P. 427. Glencoe, Ill.: Free Press, 1956.

# 49

# Test Anxiety, General Anxiety, and Intellectual Performance — *Irwin G. Sarason*

*The author reports a study evaluating the role of a general and of a specific (i.e., test) anxiety factor in academic achievement. Relationships of test anxiety and general anxiety questionnaires both to entrance examinations and to grade-point averages among liberal arts undergraduate students were examined. Although the relationships discovered were not of a sufficient magnitude to warrant accurate predictions, general anxiety scores correlated positively with grade-point averages, while test anxiety scores correlated negatively with measures of academic achievement. The importance of specific personality variables in academic achievement is apparent from the results of this study.*

PSYCHOLOGISTS CONCERNED with the problems of the prediction of the future performance of individuals have, in recent years, become increasingly explicit in acknowledging the necessity of taking personality factors into account in predicting future achievement (*9, 11, 14*). Along with this growing interest in the problem of the influence of personality variables, there has been an increase in empirical studies of the relationships between scores on paper and pencil personality tests and measures of achievement such as grade point averages (*2, 6, 9, 11, 12*).

Perhaps the personality characteristic most studied in recent years has been anxiety (as defined by score on tests like the Taylor Anxiety Scale). Since this variable has been shown to influence performance in numerous

Reprinted from *Journal of Consulting Psychology*, 1957, **21**, 485–490, by permission of the author and the American Psychological Association.

laboratory situations (*1, 4, 7, 8, 10, 13*), it is reasonable that there would be interest in extending our knowledge of its effects to "real life" behavior such as academic achievement. With respect to the Taylor scale, the results thus far have been disappointing. The most reliable studies reported in the literature indicate that level of anxiety has no demonstrable effect on academic achievement (*9, 12*). This is a rather surprising result in view of the frequent clinical observation that high anxiety leads to a breakdown or decline in achievement.

In light of this *apparent* failure to test-defined anxiety scores to predict achievement, it would seem fruitful to re-examine our notions as to what anxiety scales are measuring. In this regard, it seems that those of us who have worked with tests like the Taylor scale have ignored one very important observation made by clinicians and laymen. That is, that people are not anxious every minute of the day and that often we can specify the conditions which will lead to an increase in anxiety in the individual. Perhaps what we need are not general anxiety scales oriented towards the kinds of anxiety responses (e.g., sweating, awareness of an increase in tension, etc.) which an individual will admit to but, rather, tests designed to assess the specific conditions under which anxiety is aroused—or, of course, perhaps we need some combination of both.

The present study was designed to evaluate the role of anxiety in academic achievement when anxiety is defined as a general characteristic and also as one specific to a particular situation. Scores on two questionnaires devised by S. Sarason (*3, 5*), a general anxiety scale and a test anxiety scale, were used to select extreme anxiety groups. College entrance examination scores and grade point averages were used as response measures. It was expected that the test anxiety scale might be of more predictive utility in this situation than the general anxiety scale because of the test anxiety scale's more intensive assessment of anxiety responses and their antecedents in a specific situation, the testing situation. Evidence presented by Sarason and Mandler (*11*) and Gordon (*2*) seems to indicate that this focusing on the specifics of the test situation can lead to some increase in the ability to predict such things as achievement in academic situations. However, a direct comparison of the predictive utility of the two kinds of anxiety scales (general and test) has as yet not been attempted. In the present study, it was possible to make this kind of comparison.

A word should be said at this point of the way in which we shall use the term anxiety and the kinds of predictions this definition leads to. Our knowledge of the role of anxiety on behavior is at present quite meager, even in relatively simple experimental situations. There is, however, a growing body of evidence which can be interpreted as suggesting that individuals obtaining high scores on anxiety questionnaires differ from other subjects (Ss) in the extent to which their performance is disrupted under conditions of stress (*4, 5, 8, 10, 11*). It might further be inferred that an individual's performance is disrupted to the extent that he brings to novel situations anticipations of failure, rejection, and inability to cope with the requirements of the situation. If this were so, high anxious Ss would be regarded as emitting these interfering task-irrelevant responses (e.g., self-verbalizing, "I can't pass this test") to a greater extent than do other Ss in the anxiety score distribution. In the case where certain Ss admit to these interfering responses we would expect a lower level of performance for them than for other Ss in the score distribution. While the relationships existing between general anxiety (i.e., anxiety experienced in a wide variety of situations) and test anxiety (anxiety specifically in testing situations) are as yet far from clear, we would expect that Ss with high test anxiety scores would do relatively more poorly than low test anxious Ss in their performance on important entrance examinations for which they could not prepare. Unless general and test anxiety were very highly correlated, we would not necessarily expect individuals admitting to anxiety in a variety of situations to be anxious in a testing situation. In any event, we would expect smaller differences on entrance examinations for extreme general than for extreme test anxiety groups.

A further prediction was made with respect to the grade point averages of extreme test anxiety groups. On the assumption that high anxiety leads to performance decrements in novel situations for which preparation is not possible, it was expected that high and low test anxious groups would not differ in grade point average. This kind of result was expected on the basis that in a particular classroom situation the deleterious effects of the novelty of test situations would decrease by virtue of the increase in familiarity with the teacher, classroom, etc., over a period of a semester. Another possibility leading to the same prediction is that high test anxious Ss' anxiety may not be reduced during a course but that they may have sufficient time over the period of a semester to overlearn course material.

## Method

The Ss were 305 Yale University liberal arts undergraduate students. Most of these Ss were administered the Test Anxiety (TA) questionnaire and the General Anxiety (GA) questionnaire in the fall of 1953. In some cases, only TA scores were obtained. In all cases, the anxiety questionnaires were administered during introductory psychology class meetings. At the time of taking these questionnaires, the great majority of Ss were sophomores or juniors. It is more important to keep this fact in mind in interpreting the present results. Although the anxiety scales used are acceptably stable over time, we had no was compared with the lower 7% (L1) of the TA distribution, the low TA group was significantly superior ($p<.01$) to the high TA group on both SAT and MAT. When Ss scoring on TA between the 8th and 17th percentiles (L2) were compared with Ss scoring between the 82nd and 93rd percentiles (H2), no significant differences were obtained. It thus appears that the TA questionnaire may be sufficiently sensitive only in selecting from a large group of individuals a small subgroup of Ss who perform in the manner predicted in this study. It is interesting that this result is quite consistent with the findings in other studies that anxiety questionnaires divide Ss into two groups: a small high anxious group and a group composed of Ss in the rest of the anxiety score distribution. In this regard, it is relevant to note that the two low TA groups (L1 and L2) did not differ significantly on either of the two entrance examination measures, but that the H1 TA group was significantly inferior ($p<.05$) to the H2 TA group, and that this latter group did not differ significantly from the low TA groups. When groups of high and low scores on the GA questionnaire were compared with respect to SAT and MAT performance no significant differences were obtained.

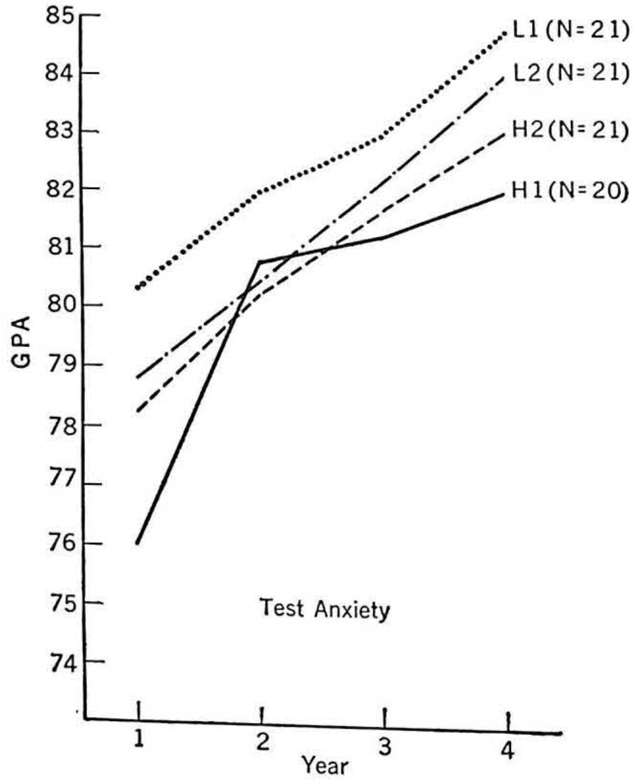


FIG. 1. *GPA as a function of years in college for four TA groups.*

In view of the correlations between TA scores and grade point averages presented in Table 1, one might expect, for the first two years, to find a superiority of low to high TA groups followed by no difference in the third and fourth years. Fig. 1 presents the curves of the H1, H2, L1, and L2 TA groups. These curves represent grade point averages as a function of years in college for Ss scoring above the 91st percentiles (H1), between the 82nd and 91st percentiles (H2), below the 9th percentile (L1), and between the 9th and 18th percentiles (L2) for the 227 Ss on whom grade point averages were available for four years. There are three observations to be made concerning Fig. 1. One aspect of it which is striking is the clearcut superiority of the L1 group over the L2 as well as over the two high TA groups. This superiority of the L1 TA group was found to be statistically significant.

Another interesting finding represented in Fig. 1 is the general similarity in the curve of the L2 group to that of the H2 group. It had been expected that the LA and HA groups would not differ significantly when grade point average was used as the measure of academic achievement. Thus, to the extent that the H2 group can be considered a high anxious group, the similarity in the L2 curve to the H1 and H2 curves is as was expected. Clearly, however, the superiority of the L1 group to the other three groups was not expected. At this point one can only conjecture as to the meaning of the superiority of this group. It is possible that a test-taking attitude (e.g., that tapped by the MMPI K scale) peculiar to L1 Ss may have contributed to their superiority in grade point average. No measure of test-taking attitude was available for the Ss used in this study. It would seem to be of interest to take this variable into account in future research.

In addition to the unexpected superiority of the L1 group to all other groups whose performance is depicted in Fig. 1 is the marked difference in form of the H1 group curve from the other three. Most striking about this group's performance is the great improvement in their performance in the second year following an extremely poor first-year showing. It had been expected that high anxious Ss would perform as well as low anxious Ss in course work on the assumption that during a semester the high anxious individual can increase his familiarity with the task, the instructor, etc., as well as possibly overlearn course material. The curves in Fig. 1 suggest that the process of acclimatization for high anxious Ss may be considerably slower than we had expected.

In order to follow up the unexpected significant positive correlations obtained between GA scores and grade point averages, two extreme high and two extreme low GA anxiety groups were compared in terms of the four yearly grade point averages obtained. Fig. 2 presents curves for the groups of Ss scoring in the upper 8% of the GA score distribution (H1), the lower 8% (L1), Ss scoring in the upper 9th to 18th percentile range, and Ss in the

82nd to 91st percentile range (H2) of the GA score distribution. A large, highly significant, superiority of high to low GA Ss is clearly in evidence throughout the four-year period. This result is not in accord with the predictions made prior to carrying out the study and, together with the surprising superiority of the L1 TA group to the other TA groups studied, poses important problems for future research. At present, it is unclear what individual difference variables are contributing to the superiority in achievement of both individuals who admit to virtually no anxiety in test situations and individuals who admit to great anxiety in a wide variety of situations.

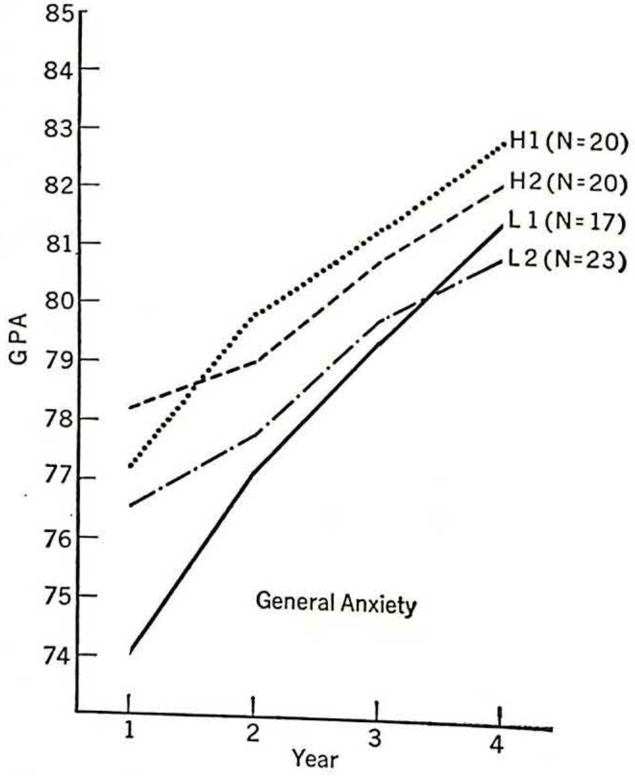


FIG. 2. *GPA as a function of years in college for four GA groups.*

It is clear, however, that it is important in discussing the effects of anxiety on performance to specify the manner in which anxiety is measured (i.e., by means of which instrument). The results of the present study reveal the importance of establishing the specific situations in which an individual experiences anxiety if one is interested in predicting his future performance in specific situations.

## Summary

1. The relationships of anxiety as measured by the Test Anxiety (TA) and General Anxiety (GA) questionnaires to entrance examinations and grade point averages were studied.

2. TA scores tended to correlate negatively with measures of academic achievement, although with increase in number of years in college the negative correlation disappeared. GA scores failed to correlate significantly with entrance examination scores, but tended to correlate positively with grade point averages.

3. In studying extreme anxiety groups, high TA Ss performed at a significantly lower level than did low TA Ss. In addition, significant differences were found within each of the low and high test anxious groups. Thus, on entrance examinations the most extreme test anxiety group performed at a significantly lower level than did a group of high anxious Ss with less extreme TA scores, and for course grades, the Ss with the lowest TA scores performed at a considerably higher level than did low anxious Ss with less extreme scores.

4. The results demonstrated that relationships between anxiety and achievement variables depend to an important extent on the nature of the instrument employed to measure anxiety.

## REFERENCES

1. Farber, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, 52, 311–327.
2. Gordon, E. M. Self-esteem, need for achievement, test anxiety, and performance. Unpublished doctoral dissertation, Yale Univer., 1954.
3. Gordon, E. M., & Sarason, S. B. The relationships between "test anxiety" and "other anxieties." *J. Pers.*, 1955, 23, 317–323.
4. Lucas, J. The interactive effects of anxiety, failure, and intraserial duplication. *Amer. J. Psychol.*, 1952, 65, 59–66.
5. Mandler, G., & Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, 47, 166–173.
6. Matarazzo, J. O., Ulett, G. A., Guze, S. B., & Saslow, G. The relationship between anxiety level and several measures of intelligence. *J. consult. Psychol.*, 1954, 18, 201–205.
7. Montague, E. K. The role of anxiety in serial learning. *J. exp. Psychol.*, 1953, 45, 91–96.
8. Sarason, I. G. Effect of anxiety, motivational instructions, and failure on serial learning. *J. exp. Psychol.*, 1956, 51, 253–260.

9. Sarason, I. G. The relationship of anxiety and "lack of defensiveness" to intellectual performance. *J. consult. Psychol.*, 1956, **20,** 220–222.
10. Sarason, I. G. Effect of anxiety and two kinds of motivating instructions on verbal learning. *J. abnorm. soc. Psychol.*, 1957, **54,** 166–171.
11. Sarason, S. B., & Mandler, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47,** 810–817.
12. Schulz, R. E., & Calvin, A. D. A failure to replicate the finding of a negative correlation between manifest anxiety and ACE scores. *J. consult. Psychol.*, 1955, **19,** 223–224.
13. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48,** 285–290.
14. Travers, R. M. W. Individual differences. *Annu. Rev. Psychol.*, 1955, **6,** 137–160.

# Index

Abilities, increases in from adolescence to adulthood, 73–76
Academic achievement, and adolescent subculture, 135–147
  and anxiety, 578
  assessment of, 536–548
  defined, 538
  human side of tests, 537–539
  and IQ, 140–143
Acceleration, 103, 104, 118–127
Adaptive flexibility, 55
Adolescence, and mental growth, 33
Adolescent development, *see* Concept attainment
Advance organizers, and learning, 420–427
Age, and ability to learn, 34
  influences on behavior problems, 182–183
Age grouping, 106
Allometric chart, 570, 571, 572
Allport-Vernon-Lindzey Study of Values, 315
Analysis of covariance, 6, 170, 203
Analysis of variance, 6, 202, 204, 253, 410, 416
Anxiety, 191, 338, 342
  and intellectual performance, 577–586
  intelligence and achievement, in children, 83–89
Arithmetic achievement, and anxiety, 88, 574, 575
Armstrong, C. M., 525; article, 567–576
Associational fluency, 54
Assumption, 287, 289–292, 297
  concluding use, 290
  deprecatory use, 290
  empirical necessity, 291, 292
  gap-filler, 290–292
  logical necessity, 292
  premise use, 290
  presupposition use, 292
Attitude change, 348–354
Attitude organization, 330–347
  competition, 336–339, 347
  discussion and conclusions, 346–347
  docility, 339–342, 347
  witch-hunt syndrome, 331–336
Auditory memory, 53
Ausubel, David, 2, 3, 394; article, 420–427; cited, 397, 432; quoted, 396
Average, composite, formula for, 563–564

Bartlett, homogeneity of variance, 409
Behavior, competence-oriented, 244
  during problem solving, 279–285
Behavior (*Continued*)
  prediction and control of, 12–17; implications, 17–18
Behavior problems, of children, 172–189
Behavior technology, goals of, 15–16
Behavioral content, 49–50
Behavioral sciences, advances in, 11–17
Behavioristic theory, 425
Behaviors, rewarded, and learning, 27
Beilin, Harry, 150; article, 172–189; cited, 180
Biggs, J. B., 228; article, 348–354
Biserial correlation, 4, 78, 79
Block-time classes, 511–523
Boehm, Leonore, 44; article, 90–102
Bogardus scale, 349, 350
Bradway, Katherine P., 44; article, 65–82; cited, 66, 67, 76, 79
Branching, 472
Bruner, Jerome, 234; article, 236–248; cited, 321, 396

California Achievement Test, 364, 370
California Mental Maturity Test, 141 n., 142 n., 365, 553
California Personality Test, 358, 359
California Psychological Inventory (CPI), 312, 316
Camouflaged Words test, 58
Central tendency, measures of, 1
Chaining, 262, 274
Check, J., article, 428–434; cited, 280, 394
Child, socialization of, 185
Childhood, and mental growth, 33
Children, transfer and retention, 428–434
Children's Manifest Anxiety Scale, 84, 89
Children's self concepts, *see* Self-concepts
Christensen, C. M., 151; article, 199–206; cited, 153
Classes, block-time, 511–523
Coefficient, correlation, 3
Cognition, factors of, 48
Cognitive abilities, 50–53
Cognitive-effective scale, 201, 202
Cognitive-need, 200
Cognitive structure, 420
Cohesiveness, and teacher behavior, 216–217
Coladarci, Arthur, 10; article 38–42
Coleman, James, 45; article, 135–147
College success, and IQ, 108
Colville, Frances, 229; article, 386–392
Competence motivation, 243
Competition, 140
  and attitude organization, 336–339, 347

Competition (*Continued*)
  and group relations, 35
  motivational factors in, 405–419
Composite average, formula for, 563 564
Concept attainment, and adolescent development, 249–257
  efficiency in, 249, 250
  perceptual error, 250
  strategy in, 249, 250
Concept formation, and teaching methods, 30
Concept learning, and problem solving, 263–276
Conditioned response, 244
Conformity, 210, 221, 222
  and the child, 241–242
  and mental growth, 34
  pressures for, 12
  and teaching methods, 29
Constant difference rule, 397, 399
Contents, and structure of intellect, 49–50
Contingencies of reinforcement, 226
Contingency table, 6
Control, of behavior, 12–17
  of human behavior, concept, 23–24
Convergent-production abilities, 56–58
Convergent thinking, 48
Cooperation, motivational factors in, 405–419
Cooperative Elementary Algebra Test, 358
Cooperative English Test, 167
Cooperative Geometry Test, 358
Cooperative Intermediate Algebra Test, 358
Cooperative Mathematics Test, 167
Cooperative Reading Test, 358
Cooperative Social Studies Test, 167
Core program development, 511–523
  orientation and pre-planning, 514–515
  support for, 520–521
Correlation, biserial, 4
  Pearson product-moment, *r*, 3, 4
  rank order, *rho*, 4
  symbol, 7
Correlation coefficient, 3–4, 368, 369
Correlation value and averaging, concept and symbol, 7
Creative talent, 305–323
  development of potential, 317–320
  family relationships, 318–319
  interests and values, 315–317
  introversion and extraversion, 315
  nature of, 310–320
  nurture of, 320–322
  openness to experience, 313
  study of architects, 308–320
Creative thinking abilities, measurement of, 121
Creativity, artistic, 307, 308
  related to intelligence, 311, 324–326
  scientific, 307–308
Critical thinking, definition, 286–304
  major aspects listed, 287–288
  major aspects clarified, 288–300
Curvilinear relation, 4

Deduction, 58
  rules for, 288
Delayed reinforcement, 226
Deviation IQ, 554
Direct Sentence Completion Test, 358, 359
Directed discovery, learning by, 396–404
Directed learning group, 398, 400–401, 402
Discovery, act of, 236–248
  defined, 237
  directed, learning by, 396–404
  heuristics of, 245–246
  intrinsic and extrinsic motives, 241–245
  learning benefits, 238
  and memory, 247–248
Dissatisfaction, with school, 355–363
Divergent thinking, 48
Divergent-thinking abilities, 53–56
Docility, and attitude organization, 339–342, 346, 347
Drummond, Harold, 464; article, 497–504
Duncan Range Test, 6, 253, 254
Dunnett procedure, 6
Dyer, Henry, 525; article, 536–548

Ebel, Robert, 523; article, 528–535
Education, and structure of intellect, 62–63
  philosophy of, 185
  relation to psychology and educational psychology, 10
Educational achievement, importance of measuring, 528–529
  testing of, 120
Educational objectives, attainment of, 225
Educational psychology, concept of, 38–39
  contribution of, 40–41
  relation to psychology and education, 10
  relevancy of, 38–42
Educational television, 32, 483–489
Elementary school, pupil-teacher relationships, 158, 161
Ends, and values in relation to science, 18–20
Ennis, Robert H., 234; article, 286–304
Ervin, Susan, 395; article, 443–461
Essay test, 532, 546
Ethical values, testing of, 121
Evaluation, and intellectual abilities, 48
Evaluative abilities, 58–60
Expository mode, teaching in, 238
Expressional fluency, 54–55
Expressional Fluency Test, 121
Extrapunitiveness, 362
Extraversion, and creativity, 315

Factor analysis, 4, 48–63
Family relationships, and creativity, 318–319

*F* distribution, 6
Feldhusen, John, 3, 44; article, 83–89; cited, 279, 280
Figural content, 52, 56, 59, 61
  defined, 48
  example, 51
Fisher's *t*, 6
Flanders, Ned A., 151; article, 207–224
Flexibility, adaptive, 55
  spontaneous, 54
Flexibility factors, 383
Fluency, associational, 54
  expressional, 54–55
  ideational, 54, 56
  word, 53–54
Forgetting, meaningful, 421
*F* ratio, 170, 171, 369

Gates Reading Test IQ, 446
General anxiety scale, 578, 579
Generalization, 287
  when warranted, 293–294
Gestalt, 26
Gestalt psychology, 298
Gestalt Transformation Test, 58
Getzels, Jacob, 228; article, 355–363; cited, 324
Gifted child, 356
Gifted pupils, and television, 484
Glock, Marvin, 525; article, 558–566
Goals, attitudinal, 539, 541–542
  of behavior technology, 15–16
  informational, 539, 540
  measurement of, 542
Grade equivalent score, 555
Grade placement score, 555
Group contagion variable, 209, 210, 215
Group discussion, and attitude change, 348–354
Group discussion skills, development of, 225–232
Group Tests, reliability coefficient of, 122
Growth age, concept of, 570
Guidance, and pupil achievement, 199
Guided discovery group, 398, 400–401, 402
Guilford, J. P., 44; articles, 46–64, 375–385; cited, 50 n., 121, 229, 283, 325
Guilford-Zimmerman Aptitude Survey, 51
Gump, Paul, 150; article, 190–198; cited, 191, 193

Handwriting skills, testing of, 121
Harris, Dale, 45; article, 128–134
Haslerud, G. M., 395; article, 435–442; cited, 436
Havumaki, Sulo, 151; article, 207–224
Health, and adolescents, 132, 133
Henmon-Nelson IQ test, 358
Henry, Jules, 228; article, 330–354
Heterogeneous grouping, 104–105

Home and family relations, and adolescents, 132
Homogeneous grouping, 34, 111
Horizontal process, and problem solving, 261, 275
Hoyt, Kenneth, 150; article, 164–171
Human behavior, control of, concept, 23–24
Hypothesis, 287
  when warranted, 294–298
Hypothetical mode, teaching in, 238

Ideational fluency, 54, 56
Immediate reinforcement, 226
Implications, cognition of, 52
Impulsion-factors, 379
Independence, and creativity, 322
Indexes, statistical, 1, 5
Individual development, stages and patterns of, 32–33
Individual differences, principles of, 505
  psychology of, 103–115
Insight, 30
Intellectual abilities, growth and decline, 69–76
Intellectual and affective characteristics, testing of, 121
Intellectual performance and anxiety, 577–586
Intellectual potency, 238–241
Intelligence, and achievement, 553–554
  adult, prediction of, 76–79
  assessment of, 61
  components of, 47, 48
  growth in childhood, 110
  measurement of, 79–80
  misconceptions, 554
    *See also* IQ
  related to creativity, 311
Interaction, concept of, 7
Interests, and abilities, 114
  and aptitudes, 554
Interference theory, 425
Intragroup aggression, 331, 333, 335, 346, 347
Intrinsic programming, 472
Intropunitiveness, 362
Introversion, and creativity, 315
Intuition, and creativity, 321
Invariance, concept, 95
Iowa Every Pupil Arithmetic Test, 358
Iowa Tests of Basic Skills, 201, 204
Iowa Tests of Educational Development, 553
IQ, and academic achievement, 140–143
  accuracy of, 551
  constancy of, 68–69

Jackson, Philip, 228; article, 355–363; cited, 324

Judging, and creativity, 321
Judgment tests, 59–60

Kates, Solis, 234; article, 249–257
Kendler, Howard, 7, 234; article, 258–278; cited, 259, 262, 264, 267, 268, 272, 470; quoted, 274, 275
Kendler, Tracy, 234, 268; article, 258–278; cited, 474
Kersh, Bert, 394; cited, 397, 398, 401
Klausmeier, Herbert J., 3, 44, 45, 234, 394; articles, 83–89, 118–127, 279–285, 428–434; cited, 279, 280; quoted, 119
Knowledge of results, 226
Koch, Margaret B., 151; article, 225–232
Kounin, Jacob S., 150; article, 190–198; cited, 191, 193
Kowitz, G. T., 525; article, 567–576
Kuhlmann-Anderson IQ, 119, 125

Language laboratories, 490–496
Language laboratory equipment, 491–492
Learning, and advance organizers, 420–427
  control of, 245
  drive-reduction model, 242
  higher aspects of, 540
  and transfer effects of a verbal generalization, 443–461
Learning process, discussed, 27–28
  teaching machines, 27
  threat and punishment, 28
  reinforcement, 27
  repetition, 27
  rewarded behaviors, 27
Longitudinal studies, 568, 569, 572, 575
Loree, M. Ray, 151; article, 225–232
Loughlin, Leo, 234; article, 279–285

MacKinnon, Donald, 234; article, 305–323
Mann-Whitney test, 6
Mathematical Aptitude Test, 580, 581, 582
Mathematicians, and creativity, 308
Mathematics, and children's thinking, 100
May, Frank B., article, 324–326; cited, 121, 235, 325
Mean scores, example, 6, 360
Meaningful learning, 403
Measurement, and the teacher, 528–535
  command of useful knowledge, 530
  educational test, defined, 528
  importance of measuring educational achievement, 528–529
  score reliability, 532–533
  teacher requirements, 531–532
  test preparation, 531
  test quality, 532
  test reliability, 534–535
  written tests, 530–531
Measuring instruments, reliability of, 4
Mediational hypothesis and theory, 265, 266, 267, 268, 269, 271, 272, 275
Memory, 435
  conservation of, 247–248
  and intellectual ability, 48
Memory abilities, 53
Memory-reasoning task, 405, 414, 417
Mental defectives, education of, 112
Mental growth, rate of, 33
Mental health, and adolescents, 132
  ranking of behavioral problems, 179, 180
Method, efficiency in problem solving, 280, 281, 282
Methods, *see* Teaching methods
Metropolitan Achievement Tests, 119, 120, 125
Meyers, Shirley, 395; article, 435–442
Miller, K. M., 228; article, 348–354
Millman, Jason, 525; article, 558–566
Minnesota Multiphasic Personality Inventory (MMPI), 311
Minnesota Teacher Attitude Inventory, 205
Model, mediational S-R, 258
Models, for problem solving, 260, 260–263
Mode of attack, and problem solving, 280, 281, 282
Moral qualities, and adolescents, 132
Moreno technique, 349
Morrill, Charles, 464; article, 466–482
Motivating effect, of learning by directed discovery, 396–404
Motivation, 199, 200
  discussed, 28–29
  in cheating, 28
  learning readiness, factors, 28
  rewards, 28
  and subject matter, 31
  and teaching machines, 476–477
  theory of, 242
Motivational factors, in cooperation and competition, 405–419
Motives, intrinsic and extrinsic, 241–245
Motor performance, 446, 455
Motor skills, 24
  and mechanical principles, 386–392
Motor-speed factors, 381–382
Multiple-choice test, 533, 546

Nongraded programs, 505–510
Nonpunitive teachers, 190–198
Nonreversal shift, example, 263, 264, 264 n., 265, 265 n., 266, 266 n., 267, 268, 273
Nonverbal quotient, 76
Normal-curve critical-ratio test, 6
Number concept quotient, 76

Object Improvement Test, 121
Object Uses Test, 121
Obliterative subsumption, decelerating rate of, 421
Observation statement, 287
  reliability of, 292–293

Operations, and structure of intellect, 49–50
Originality, 55, 56
  and creativity, 306–307
Otis Beta test, 253
Otis Quick Scoring Mental Ability Test, 166, 167, 325
Overachievement, 242

Pearson product-moment correlation, *r*, 3, 4, 77, 230, 580
Peer acceptance, measurement of, 121
Perceptual error, 250, 253, 254
Perceptual-motor task, 405
Perkins, Hugh, 228, 468; article, 364–374, 505–510
Permissiveness scale, 201, 202
Persistence, 240
  and problem solving, 280, 281, 282
Personality changes, 13–14
Personality variables, 577
Pertinent Questions Test, 53
Philosophy of life, and adolescents, 132
Physics, similarity to psychology, 14–15
Picture Arrangement Test, 57
Plot Titles Test, 55–56, 121, 125
Plot Questions Test, 121, 125
Prediction, of behavior, 12–17
  of future achievement, 577
Pretheoretical model, 260, 261–263
Primary Mental Abilities Test, 446
Principles, given and derived, transfer value of, 435–442
Problem identification, reduction of, 301–302
Problem solving, 246, 248, 258–278
  ability, testing of, 120–121
  behaviors during, 279–285
  concept learning and utilization, 263–276
  history of, 259
  methodological problems, 259–261
  pretheoretical model, 261–263, 274, 275
Product-moment correlation, Pearson, *r*, 3, 4, 77, 230, 580
Products, and structure of intellect, 49–50
Programs, nongraded, 505–510
Promotion policy and academic achievement, 567–576
Proprioceptive stimulation, in learning, 265
Psychegroups, 350, 351, 352, 353, 354
Psychoanalysis, 48
Psychology, advances in, 11–17
  educational, characteristics of, 1
  relation to educational psychology and education, 10
  similarity to physics, 14–15
Psychological health, in school, 355–363
Psychological theory, and structure of intellect, 60–61
Psychomotor abilities, system of, 375–385
  general implications, 385
  testing of, 121
  whole-body factors, nature of, 383–384
Psychomotor factors, 378–383
Psychotherapy, and individual freedom, 17–18, 22–23
  studies in, 13–14
Punishment, and learning, 28
Punitive teachers, 190–198
Pupil achievement, and teacher permissiveness, 199–206
Pupil behavior dimensions, 157–159
Pupil behavior index, 159–161

Q-sort, defined, 365
Q-sort instrument, reliability of, 366–368

Raven Matrices Test, 110
Reading, and learning, 30
Reading achievement, 574–575
Reasoning, 73, 74, 75
  categorical, 288
  general, defined, 52
  if-then, 288
  syllogistic, 288
Recall, spontaneous, 247
Redefinition abilities, defined, 57
Reed's Sentence Completion Test, 365
Reinforced competition, 337
Reinforcement, 16, 475
  immediate, 469
  intermittent, 264, 266–267
  and learning, 27
  secondary, 242
    *See also* Reward
Reliability, of measuring instruments, 4
Reliability coefficient, 4
  of group tests, 122
Repetition, and learning, 27
Response mode, 473–476
Retention, 30–31, 339
  and advance organizers, 420–427
  test of, 421–422, 427
  and transfer in children, 428–434
Retention scores, standard deviation of, 2
Retroactive inhibition, 401
Reversal shift, example, 263, 264, 264 n., 265, 265 n., 266, 266 n., 267, 268, 270, 272, 273
Reward, 226, 264
  and learning, 27
  and motivation, 28
    *See also* Reinforcement
Ripple, Richard E., 45; articles, 119–127, 324–326
Rogers, Carl, 10; article, 11–25
Role theory, 172
Rorschach Test, 359

Rote learning, 101, 399, 400–401, 402, 426, 427
  and creativity, 321
Rote memory, 73, 74, 75
$r$, Pearson product-moment correlation, 3, 4, 77, 230, 580
Ruleg, 472
Russell Sage Social Relations Test, 547
Ryans, David, 4, 150; article, 152–163; cited 153, 154, 158, 160

Sarason, S. B., article, 576–586; cited, 84, 87, 526
Sarason Test Anxiety Scale, 84
Scale, characteristics of, 202
  cognitive-affective, 201, 202
  Permissiveness, 201, 202
  of problem behaviors, 176
  Warmth, 201, 202
Scholastic Aptitude Test, 580, 581, 582
School, dissatisfaction with, 355–363
  nongraded, characteristics, 506
School and learning, testing of attitudes, 120
School marks, assignment of, 558–566
  basic concept, 559
  composite average, 563–564
  composite score, 560
  final grade assignment, 564–565
  stanines, 561–563
Science, ends and values in relation to, 18–20
"Science of Learning and the Art of Teaching, The," 468
Scientific creativity, 307–308
Score meanings, misconceptions, 554–555
Scrambled Words Test, 50–51
Second signalling system, 244
Secondary reinforcement, 242
Secondary school pupil-teacher relationships, 158, 159
Self-concept, defined, 365
Self-concepts, change in children's, 364–374
Self-discovery, learning by, 403
Self-ideal, congruency, 365, 368, 370, 371, 372, 373, 374
Semantic content, 53, 57, 58, 59, 61
  defined, 48
  example, 51
Sensitivity to problems, 60
Sex differences, and adolescent life problems, 128–134
  and behavior problems, 181–183
  and IQ, 71–73
Shane, Milton L. and Helen Lacy, 464; article, 490–496
Shaw, Marvin, 394; article, 405–419
Single Unit S-R Theory, 258, 264, 265, 266, 267, 268, 271, 272, 275
Social class grouping, 105
Social distance, 352
Social intelligence, 50, 61–62
Social interaction, benefits, 35
Society, closed, 24
  controlled, 16–18
  open, 24
Socioeconomic classes, 36
Socioeconomic status, and behavior problems, 183
Sociograms, 350
Sociogroups, 350, 351, 352, 353, 354
Spatial Orientation test, 51
Spearman-Brown prophecy formula, 390
Spearman-Brown reliability formula, 534
Spontaneous flexibility, 54
Spontaneous verbalization, transfer effects of, 456–457
SRA Primary Mental Abilities Test, 141 n.
S-R associations, 261
S-R learning, pretheoretical model, 260
S-R principles, 263
S-R relationships, 265
Standardized tests, 32, 542, 543, 549, 550, 555
Stanford-Binet, 64, 65, 66, 67, 68, 70, 71, 74, 75, 76, 77, 78, 79, 80, 81, 107, 437, 554
Stanines defined, 561–563
Statistical design, 170
Statistical significance, analysis of covariance, 6
  analysis of variance, 6
  chi-square, 6
  Duncan range test, 6
  Dunnett procedure, 6
  $F$ test, 6
  Fisher's $t$, 6
  Mann-Whitney test, 6
  normal-curve critical ratio test, 6
  *vs.* practical significance, 5
Statistical symbols and abbreviations, 7
Stern, concept of IQ, 66
Stimulus, and motivation, 242
  and response events, 272
Stimulus pattern, 273
Stimulus pre-differentiation, 444
Stimulus substitution, 244
Stimulus support, 469
Strategy, event-matching, 239
  features of, 240
Strength factors, 379
Strong Vocational Interest Blank, 312, 315
Structure of intellect, 46–63, 375
Structure of knowledge, 321
Studies, unified, 513
Study habits, and adolescents, 132
Subjective choice, in science, 18–20
Subliminal stimuli, 2–3
Substitute gratification, 341
Subsumption theory, 420, 421

Syllogistic-type tests, 59
Symbolic content, 57, 58, 59, 61
  defined, 48
  example, 51
Symbolic units, ability to cognize, 50–51
Symbols, *see* Statistical symbols and abbreviations

Taylor Anxiety Scale, 577, 578
Teacher, and educational television, 488–489
  function of, 185
  and measurement, 528–535
  role of, 152
Teacher attitudes, 175
Teacher behavior, 156–157, 160, 161, 162
Teacher characteristics, 152–163
Teacher description of problem behavior, 176
Teacher influence, 207–227
Teacher rating scale, 122, 123
Teacher requirements, for testing, 531–532
Teachers' ranking of behavioral problems, 179, 180
Teaching, by example, 435
  expository mode, 238
  hypothetical mode, 238
  by stating the principle, 435
Teaching machines, 27, 226, 466–482
  current trends, 467–471
  definition, 471
  general problem areas, 471–478
  knowledge of results, 476
  motivation, 476–477
  problems of application, 478–479
  programming, 471–473
  response mode, 473–476
Teaching methods, discussed, 29–31
  concept formation, 30
  criticism, 29
  discipline, 29
  excessive teacher direction, 29
  insight, 30
  participation, 29
  reading, 30
  retention, 30–31
  thought process, 30
Team teaching, 497–504
  an assessment, 499–504
  types of, 498–499
Television, *see* Educational television
Terman Concept Mastery Test, 311
Test preparation, 531
Test reliability, 534–535
Testing programs, 549–557
  confusion of norms and standards, 551–552
  determination of vocational goals, 552–553
  future success prediction, 552
  instruction effectiveness, 556
Testing programs (*Continued*)
  intelligence and achievement, 553–554
  interests and aptitudes, 554
  pupil categorization, 551
  score meanings, 554–555
Tests, educational, defined, 529
  one-tailed, 5
  ordering, 57
  quality of, 532
  score reliability, 532–533
  standardized, 32
  of statistical significance, *see* Statistical, significance, tests of
  written, 530–531
Therapist, attitudes of, 13–14
Theoretic systems, criteria for, 298–299
Theoretical distributions, of results, *see* Distributions, theoretical, of results
Theoretical value, 315
Theory, 287
  when warranted, 298–299
Thinking, convergent, 48
  divergent, 48
Thinking ability, testing of, 543–545
Thompson, Clare W., article, 65–82; cited, 44, 66, 67
Trait theory, 154
Transfer, 110, 399
  and retention in children, 428–434
  of principles, 387, 435–442
Transfer effects of learning a verbal generalization, 443–461
Transformations, defined, 52

Unified studies, 513
Unusual Details Test, 60

Value choice, and science, 18–20
Value statement, judging of, 287, 301–302
Value systems, variations in, 135
Values, and ends in relation to science, 18–20
Vanishing, 472
Variability, in test scores, 560
  measures of, 2–3
Vars, Gordon, 465; article, 511–523
Verbal behavior, 269
Verbal comprehension, 57
Verbal generalization, transfer effects of learning, 443–461
Verbal mediation, 446
Verbal performance, and motor performance, 456
Verbal quotient, 76
Verbal reinforcement, 225–232
Verbal responses, in problem solving, 270, 271
Vernon, Philip E., 44; article, 103–117
Vertical process, and problem solving, 261, 262, 275
Visual memory, 53

Visual-motor coordination, 444, 457, 459
Vocational testing, and structure of intellect, 61–62

*Walden Two*, 24
  quoted and discussed, 15–17
Warmth scale, 201, 202
Watson, Goodwin, 10; article, 26–37
Wechsler Adult Intelligence Scale, 65, 76
*What Psychology Can We Trust?*, 26 n.
Whole-body factors, nature of, 383–384
Wickman rating scale, 176
Wickman Study, 172, 173–178, 183–184, 185, 186
WISC IQ, 85, 89, 280, 282, 283, 429
Witch-hunt Syndrome, 331–336
Withall's Climate Index, 365
Withdrawal, 178, 179, 180, 181, 184, 281
Wittich, Walter, 464; article, 483–489
Womer, Frank, 525; article, 549–557
Word fluency, 53–54
Word Grouping Test, 57
Word Uses Tests, 121
Written tests, 546

Yudin, Lee, 234; article, 249–257; cited, 251, 252

Zeigarnik effect, 403

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road